Origin and evolution of pathogenic coronaviruses

Jie Cui [ID][1], Fang Li[2] and Zheng-Li Shi [ID][1]*

Abstract | Severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) are two highly transmissible and pathogenic viruses that emerged in humans at the beginning of the 21st century. Both viruses likely originated in bats, and genetically diverse coronaviruses that are related to SARS-CoV and MERS-CoV were discovered in bats worldwide. In this Review, we summarize the current knowledge on the origin and evolution of these two pathogenic coronaviruses and discuss their receptor usage; we also highlight the diversity and potential of spillover of bat-borne coronaviruses, as evidenced by the recent spillover of swine acute diarrhoea syndrome coronavirus (SADS-CoV) to pigs.

Severe acute respiratory syndrome
A serious form of pneumonia that is characterized by diffuse alveolar damage and that has the potential to progress to acute respiratory distress.

Type II pneumocytes
Epithelial cells that line the lung alveoli; type II cells are round and produce surfactants to lower the surface tension of water and allow the membrane to separate, thereby increasing the capability to exchange gases.

[1]*CAS Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, Hubei, China.*

[2]*Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, USA.*

*e-mail: zlshi@wh.iov.cn

Coronaviruses cause respiratory and intestinal infections in animals and humans[1]. They were not considered to be highly pathogenic to humans until the outbreak of severe acute respiratory syndrome (SARS) in 2002 and 2003 in Guangdong province, China[2–5], as the coronaviruses that circulated before that time in humans mostly caused mild infections in immunocompetent people. Ten years after SARS, another highly pathogenic coronavirus, Middle East respiratory syndrome coronavirus (MERS-CoV) emerged in Middle Eastern countries[6]. SARS coronavirus (SARS-CoV) uses angiotensin-converting enzyme 2 (ACE2) as a receptor and primarily infects ciliated bronchial epithelial cells and type II pneumocytes[7,8], whereas MERS-CoV uses dipeptidyl peptidase 4 (DPP4; also known as CD26) as a receptor and infects unciliated bronchial epithelial cells and type II pneumocytes[9–11]. SARS-CoV and MERS-CoV were transmitted directly to humans from market civets and dromedary camels, respectively[12–14], and both viruses are thought to have originated in bats[15–21]. Extensive studies of these two important coronaviruses have not only led to a better understanding of coronavirus biology but have also been driving coronavirus discovery in bats globally[21–31]. In this Review, we focus on the origin and evolution of SARS-CoV and MERS-CoV. Specifically, we emphasize the ecological distribution, genetic diversity, interspecies transmission and potential for pathogenesis of SARS-related coronaviruses (SARSr-CoVs) and MERS-related coronaviruses (MERSr-CoVs) found in bats, as this information can help prepare countermeasures against future spillover and pathogenic infections in humans with novel coronaviruses.

Coronavirus diversity
Coronaviruses are members of the subfamily *Coronavirinae* in the family *Coronaviridae* and the order *Nidovirales* (International Committee on Taxonomy of Viruses). This subfamily consists of four genera — *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* — on the basis of their phylogenetic relationships and genomic structures (FIG. 1). The alphacoronaviruses and betacoronaviruses infect only mammals. The gammacoronaviruses and deltacoronaviruses infect birds, but some of them can also infect mammals[24]. Alphacoronaviruses and betacoronaviruses usually cause respiratory illness in humans and gastroenteritis in animals. The two highly pathogenic viruses, SARS-CoV and MERS-CoV, cause severe respiratory syndrome in humans, and the other four human coronaviruses (HCoV-NL63, HCoV-229E, HCoV-OC43 and HKU1) induce only mild upper respiratory diseases in immunocompetent hosts, although some of them can cause severe infections in infants, young children and elderly individuals[1,28,29]. Alphacoronaviruses and betacoronaviruses can pose a heavy disease burden on livestock; these viruses include porcine transmissible gastroenteritis virus[32], porcine enteric diarrhoea virus (PEDV)[33] and the recently emerged swine acute diarrhoea syndrome coronavirus (SADS-CoV)[34]. On the basis of current sequence databases, all human coronaviruses have animal origins: SARS-CoV, MERS-CoV, HCoV-NL63 and HCoV-229E are considered to have originated in bats; HCoV-OC43 and HKU1 likely originated from rodents[28,29]. Domestic animals may have important roles as intermediate hosts that enable virus transmission from natural hosts to humans. In addition, domestic animals themselves can suffer disease caused by bat-borne or closely related coronaviruses: genomic sequences highly similar to PEDV were detected in bats[35–38], and SADS-CoV is a recent spillover from bats to pigs[34] (FIG. 2). Currently, 7 of 11 ICTV-assigned *Alphacoronavirus* species and 4 of 9 *Betacoronavirus* species were identified only in bats (FIG. 3). Thus, bats are
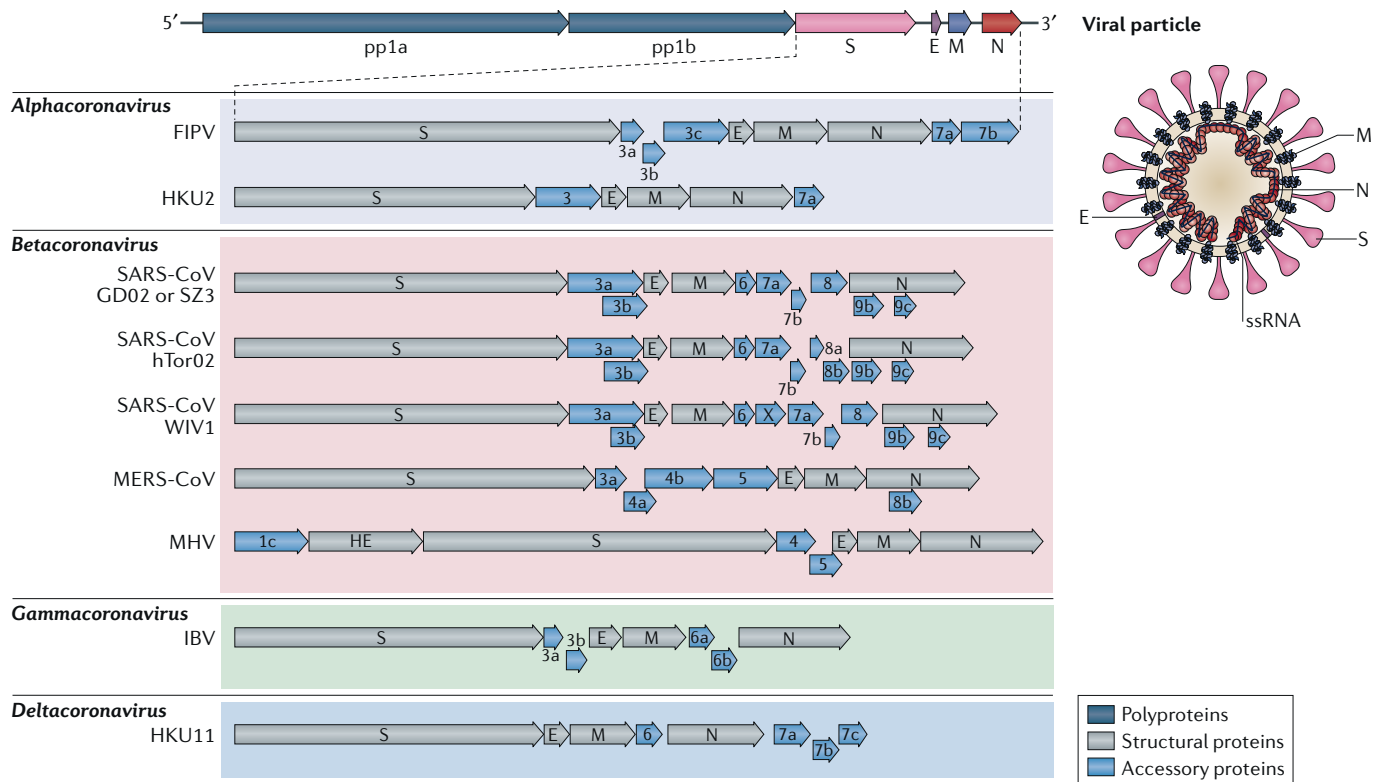
Fig. 1 | **The genomes, genes and proteins of different coronaviruses.** Coronaviruses form enveloped and spherical particles of 100–160 nm in diameter. They contain a positive-sense, single-stranded RNA (ssRNA) genome of 27–32 kb in size. The 5'-terminal two-thirds of the genome encodes a polyprotein, pp1ab, which is further cleaved into 16 non-structural proteins that are involved in genome transcription and replication. The 3' terminus encodes structural proteins, including envelope glycoproteins spike (S), envelope (E), membrane (M) and nucleocapsid (N). In addition to the genes encoding structural proteins, there are accessory genes that are species-specific and dispensable for virus replication. Here, we compare prototypical and representative strains of four coronavirus genera: feline infectious peritonitis virus (FIPV), *Rhinolophus* bat coronavirus HKU2, severe acute respiratory syndrome coronavirus (SARS-CoV) strains GD02 and SZ3 from humans infected during the early phase of the SARS epidemic and from civets, respectively, SARS-CoV strain hTor02 from humans infected during the middle and late phases of the SARS epidemic, bat SARS-related coronavirus (SARSr-CoV) strain WIV1, Middle East respiratory syndrome coronavirus (MERS-CoV), mouse hepatitis virus (MHV), infectious bronchitis virus (IBV) and bulbul coronavirus HKU11.

likely the major natural reservoirs of alphacoronaviruses and betacoronaviruses[24].

Animal origin and evolution of SARS-CoV

At the beginning of the SARS epidemic, almost all early index patients had animal exposure before developing disease. After the causative agent of SARS was identified, SARS-CoV and/or anti-SARS-CoV antibodies were found in masked palm civets (*Paguma larvata*) and animal handlers in a market place[12,16,39–42]. However, later, wide-reaching investigations of farmed and wild-caught civets revealed that the SARS-CoV strains found in market civets were transmitted to them by other animals[16,39]. In 2005, two teams independently reported the discovery of novel coronaviruses related to human SARS-CoV, which were named SARS-CoV-related viruses or SARS-like coronaviruses, in horseshoe bats (genus *Rhinolophus*)[15,43]. These discoveries suggested that bats may be the natural hosts for SARS-CoV and that civets were only intermediate hosts. Subsequently, many coronaviruses phylogenetically related to SARS-CoV (SARSr-CoVs) were discovered in bats from different

provinces in China and also from European, African and Southeast Asian countries[15,20,38,43–54] (FIG. 4; Supplementary Fig. S1a). According to the ICTV criteria, only the strains found in *Rhinolophus* bats in European countries, Southeast Asian countries and China are SARSr-CoV variants. Those from *Hipposideros* bats in Africa are less closely related to SARS-CoV and should be classified as a new coronavirus species[54]. These data indicate that SARSr-CoVs have wide geographical spread and might have been prevalent in bats for a very long time. A 5-year longitudinal study revealed the coexistence of highly diverse SARSr-CoVs in bat populations in one cave of Yunnan province, China[18,20,55]. This location is a diversity hot spot, and the SARSr-CoVs in this location contain all the genetic diversity found in other locations of China. Furthermore, the viral strains that exist in this one location contain all genetic elements that are needed to form SARS-CoV (FIG. 5). As no direct progenitor of SARS-CoV was found in bat populations despite 15 years of searching and as RNA recombination is frequent within coronaviruses[56], it is highly likely that SARS-CoV newly emerged through recombination of

bat SARSr-CoVs in this or other yet-to-be-identified bat caves. This hypothesis is consistent with previous data showing that a direct progenitor of SARS-CoV emerged before 2002 (REFS[42,57,58]). Recombination analysis also strongly supported the hypothesis that the civet SARS-CoV strain SZ3 arose through recombination of two existing bat strains, WIV16 and Rf4092 (REF.[20]).
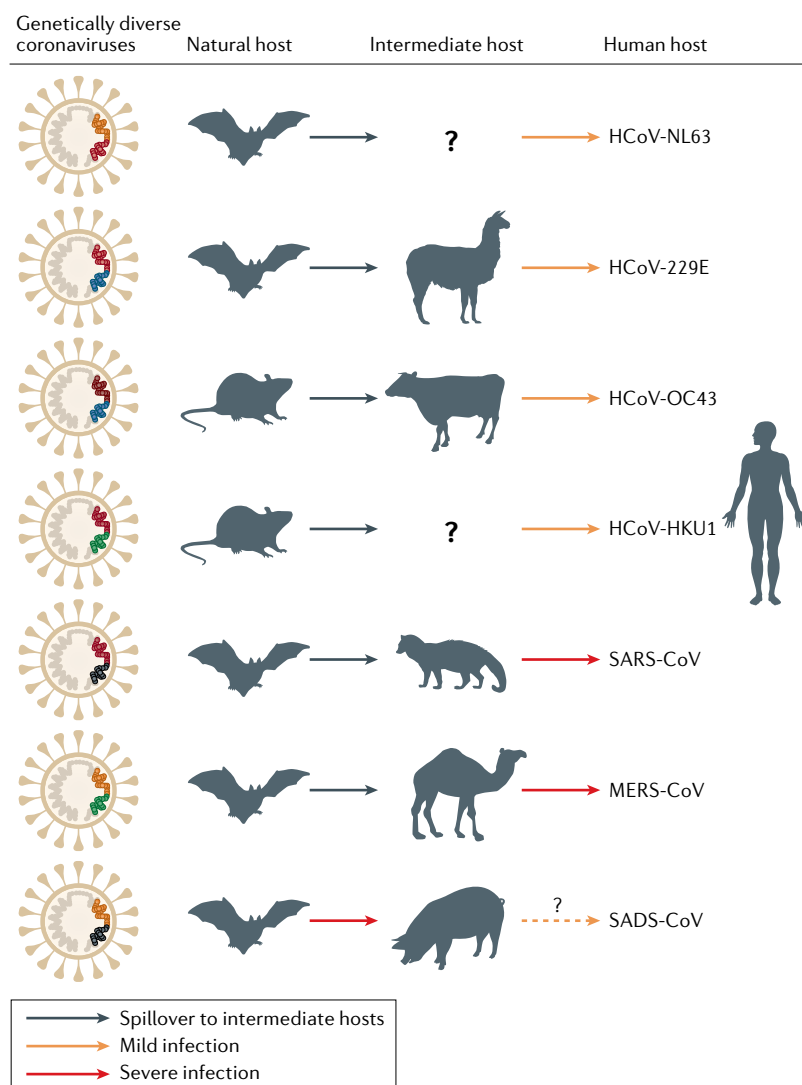

Fig. 2 | **Animal origins of human coronaviruses.** Severe acute respiratory syndrome coronavirus (SARS-CoV) is a new coronavirus that emerged through recombination of bat SARS-related coronaviruses (SARSr-CoVs)[20]. The recombined virus infected civets and humans and adapted to these hosts before causing the SARS epidemic[42,62]. Middle East respiratory syndrome coronavirus (MERS-CoV) likely spilled over from bats to dromedary camels at least 30 years ago[100] and since then has been prevalent in dromedary camels. HCoV-229E and HCoV-NL63 usually cause mild infections in immunocompetent humans. Progenitors of these viruses have recently been found in African bats[133,134], and the camelids are likely intermediate hosts of HCoV-229E[134,135]. HCoV-OC43 and HKU1, both of which are also mostly harmless in humans, likely originated in rodents. Recently, swine acute diarrhoea syndrome (SADS) emerged in piglets. This disease is caused by a novel strain of *Rhinolophus* bat coronavirus HKU2, named SADS coronavirus (SADS-CoV)[34]; there is no evidence of infection in humans. Solid arrows indicate confirmed data. Broken arrows indicate potential interspecies transmission. Black arrows indicate infection in the intermediate animals, yellow arrows indicate a mild infection in humans, and red arrows indicate a severe infection in humans or animals.

Furthermore, WIV16, the closest relative to SARS-CoV found in bats, likely arose through recombination of two other prevalent bat SARSr-CoV strains[20]. The most frequent recombination breakpoints are within the *S* gene, which encodes the spike (S) protein that contains the receptor-binding domain (RBD), and upstream of *orf8*, which encodes an accessory protein[20,58,59]. Given the prevalence and great genetic diversity of bat SARSr-CoVs, their close coexistence and the frequent recombination of the coronaviruses, it is expected that novel variants will emerge in the future[60,61]. Because there were no SARS cases in Yunnan province during the SARS outbreak, we hypothesize that the direct progenitor of SARS-CoV was produced by recombination within bats and then transmitted to farmed civets or another mammal, which then transmitted the virus to civets by faecal–oral transmission. When the virus-infected civets were transported to Guangdong market, the virus spread in market civets and acquired further mutations before spillover to humans.

Variability of SARS-CoV in humans and civets

The genome sequences of SARS-CoVs from market civets are almost identical to the genomes of human SARS-CoVs[42,62]. However, two genes show major variation. The first variable region is located in the *S* gene. The SARS-CoV S protein is functionally divided into two subunits, denoted S1 and S2, which are responsible for receptor binding and fusion with the cellular membrane, respectively[1]. S1 is further divided into the amino-terminal domain (S1-NTD) and the carboxy-terminal domain (S1-CTD). The S1-CTD functions as the RBD and is responsible for binding ACE2 and entering cells[7,63,64]. Two amino acid residues in the RBD, 479 and 487, were identified to be essential for ACE2-mediated SARS-CoV infection and critical for virus transmission from civets to humans[76,78].

The second major location of variation is the accessory gene *orf8* (FIG. 5). On the basis of SARS spread, the SARS 2002–2003 outbreak could be divided into three phases, with the early phase characterized by a limited number of localized cases, followed by a middle phase during which a superspreader event occurred in a hospital and finally the late phase of international spread[62]. The viral genomes from early-phase patients contain two genotypes of *orf8*, one with a complete *orf8* (369 nucleotides) and the other containing an 82-nucleotide deletion. By contrast, viral genomes from late-phase patients and most of the genomes from middle-phase patients contain a split *orf8* (*orf8a* and *orf8b*) owing to a 29-nucleotide deletion; two exceptions were found in middle-phase genomes, one containing an 82-nucleotide deletion in *orf8* and the other with the whole *orf8* deleted. The human isolates from 2004 and all civet SARS-CoV genomes have a complete *orf8* except one civet strain with an 82-nucleotide deletion[62]. These data indicate that *orf8* genes underwent adaptations during transmission from animals to humans during the SARS epidemic. A limited functional analysis suggested that the ORF8a protein is dispensable for SARS-CoV replication in Vero E6 cells but may have a role in modulating endoplasmic reticulum
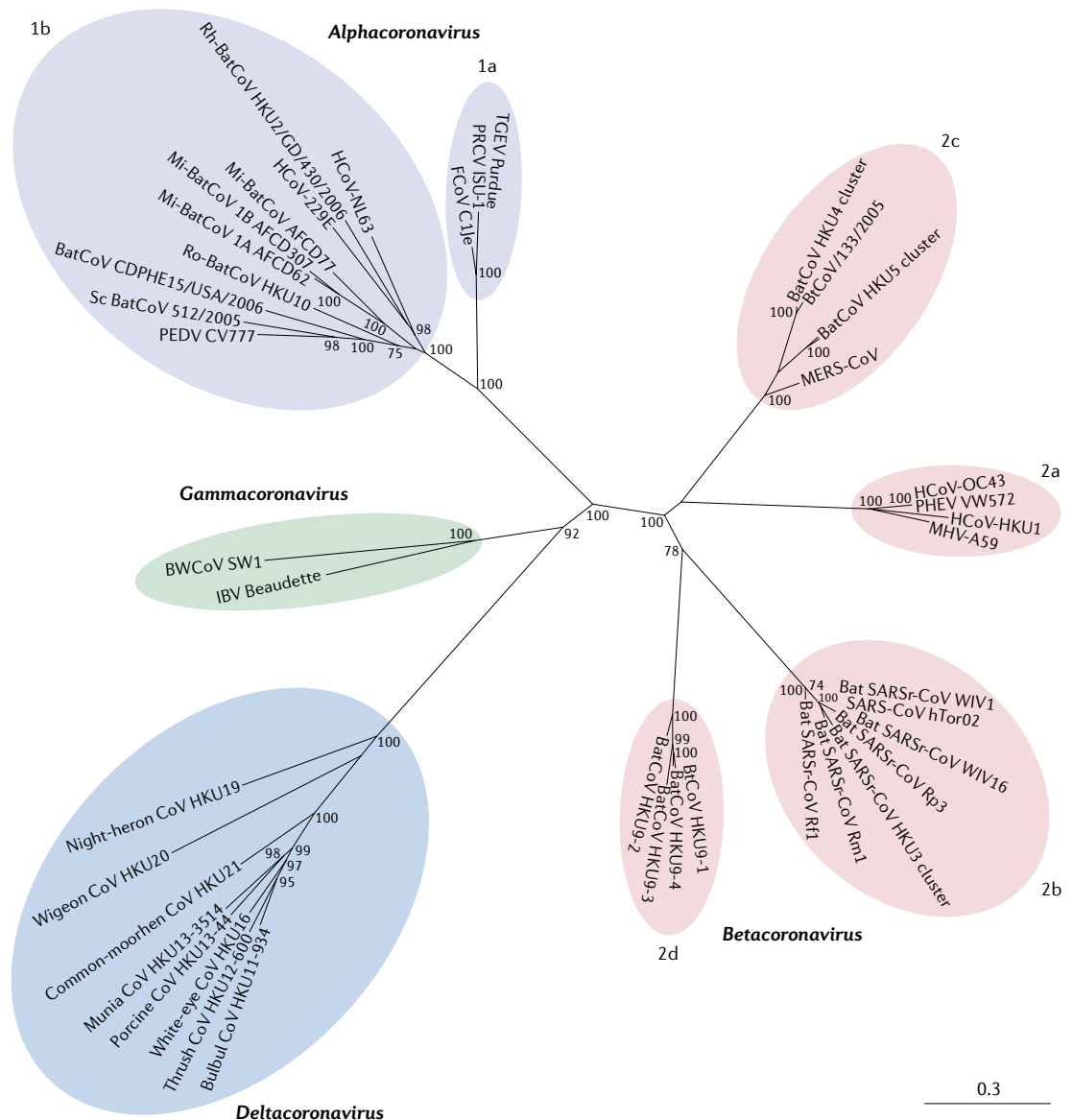
Fig. 3 | **Phylogenetic relationships in the *Coronavirinae* subfamily.** The highly human-pathogenic coronaviruses belong to the subfamily *Coronavirinae* from the family *Coronaviridae*. The viruses in this subfamily group into four genera (prototype or representative strains shown): *Alphacoronavirus* (purple), *Betacoronavirus* (pink), *Gammacoronavirus* (green) and *Deltacoronavirus* (blue). Classic subgroup clusters are labelled 1a and 1b for the alphacoronaviruses and 2a–2d for the betacoronaviruses. The tree is based on published trees of *Coronavirinae*[25,136] and reconstructed with sequences of the complete RNA-dependent RNA polymerase-coding region of the representative coronaviruses (maximum likelihood method under the GTR+I+Γ model of nucleotide substitution as implemented in PhyML, version 3.1 (REF.[137])). Only nodes with bootstrap support above 70% are shown. IBV, infectious bronchitis virus; MERS-CoV, Middle East respiratory syndrome coronavirus; MHV, mouse hepatitis virus; PEDV, porcine enteric diarrhoea virus; SARS-CoV, severe acute respiratory syndrome coronavirus; SARSr-CoV, SARS-related coronavirus.

stress, inducing apoptosis and inhibiting interferon responses in host cells[20,65–69]. Whether and how these adaptations were involved in SARS-CoV virulence are not fully clarified.

Variability of bat SARSr-CoVs
SARS-CoVs and bat SARSr-CoVs mainly vary in three regions: S, ORF8 and ORF3 (FIG. 5). Bat SARSr-CoVs share high sequence identity with SARS-CoV in the S2 region but are highly variable in the S1 region. Compared with human and civet SARS-CoV, bat SARSr-CoV

S1 can be divided into two clades: clade 1, which is found only in Yunnan province, has the same size S protein as human and civet isolates[18–20,51], whereas clade 2, which is found in many locations, has a shorter size S protein owing to deletions of 5, 12 or 13 amino acids in length[15,43–45,48,50]. Among the sequenced bat SARSr-CoVs, those with deletions in their RBDs show 78.2–80.2% amino acid sequence identity with SARS-CoV in the S protein, whereas those without deletions are much more closely related to SARS-CoV, with 90.0–97.2% amino acid sequence identity.

The second variable region is located in ORF8. Most of bat SARSr-CoVs retain an intact *orf8* (366 or 369 nucleotides) and share 47.7–100% sequence identity among themselves and 50.6–98.4% with SARS-CoV in civets and early-phase patients. A split *orf8* (364 nucleotides) owing to a 5-nucleotide deletion was found in one bat SARSr-CoV strain, similar to that of SARS-CoVs from middle-phase and late-phase patients[20]. The European bat SARSr-CoV has completely lost *orf8* (REF.[45]). These data show that the *orf8* genes in bat SARSr-CoVs are constantly evolving in their natural reservoirs. Considering the variability of *orf8* in bats, civets and humans, investigating the function of *orf8* is a priority, particularly the contribution of these different variants to viral pathogenesis.

The third variable region is in ORF3. The SARS-CoV genome encodes a 154-amino acid ORF3b, which is an interferon antagonist. Bat SARSr-CoVs and SARS-CoV are highly similar in ORF3a (96.4–98.9% amino acid identity), but bat SARSr-CoVs have different sizes of ORF3b (54–154 amino acids) (a large part of the region encoding ORF3b overlaps with the ORF3a coding region)[20,70]. ORF3b retains the anti-interferon function in some bat SARSr-CoVs but has lost the function in other bat SARSr-CoVs[70].

A novel accessory gene, named *orfx* and located between *orf6* and *orf7*, was identified in the genomes of several bat SARSr-CoVs from Yunnan province[18–20] (FIG. 5). A preliminary study indicated that ORFX is involved in an anti-interferon response[71].

## Receptor usage of SARS-CoV and SARSr-CoV

ACE2 binding is a critical determinant for the host range of SARS-CoV[72,73]. Electron microscopic studies have shown that the SARS-CoV S protein forms a clover shaped trimer, with three S1 heads and a trimeric S2 stalk[74,75]. The RBD is located on the tip of each S1 head. The RBD binds to the outer surface of ACE2, away from its zinc-chelating enzymatic site[77,141] (FIG. 6a). Different SARS-CoV strains isolated from several hosts vary in their binding affinities for human ACE2 and consequently in their infectivity of human cells[76,78] (FIG. 6b). The epidemic strain hTor02 was isolated from humans during the late phase of the outbreak in 2002–2003. It has a high affinity for human ACE2 and high infectivity in human cells, and consequently, it was transmitted efficiently between humans[62]. Strains cSz02 and cHb05 were isolated from palm civets in 2002–2003 and 2005, respectively. Both have low affinity for human ACE2 and low infectivity in human cells but have high affinity for civet ACE2 and high infectivity in civet cells[12,79]. Strain hcGd03 was isolated from both humans and palm civets in 2003–2004 and has moderate affinity for human ACE2 and moderate infectivity in human cells; it infected humans but did not transmit between humans[80]. Strain hHae08 was isolated from human cell culture and has high affinity for human ACE2 and high infectivity in human cells[81]. Understanding the molecular basis for human receptor usage by different SARS-CoV strains is crucial for understanding the cross-species transmission of SARS-CoV and for epidemiological monitoring of potential future outbreaks.

SARS-CoV mutations that affect human and civet receptor binding. Crystal structures of the SARS-CoV RBD complexed with human ACE2 revealed that the SARS-CoV RBD contains a core structure and a receptor-binding
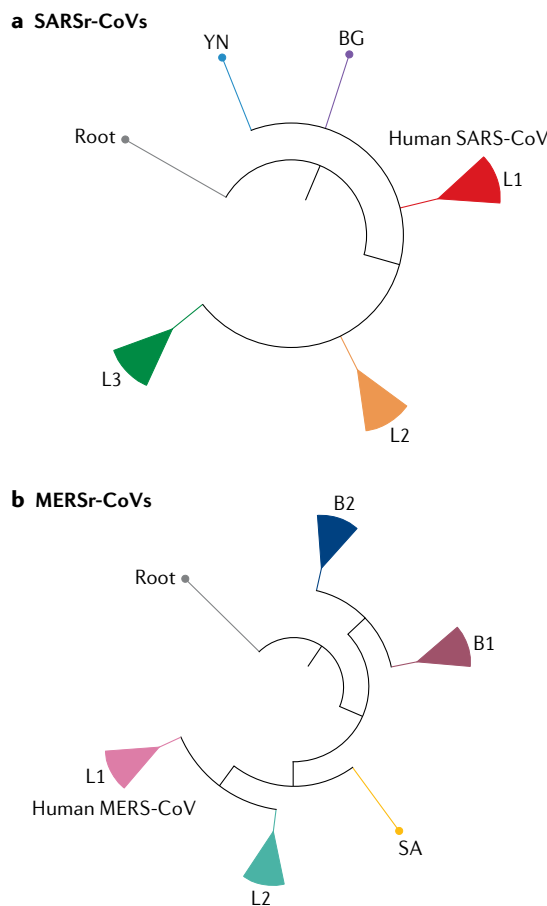
Fig. 4 | **Phylogenetic analysis of SARSr-CoVs and MERSr-CoVs. a** | The figure shows a simplified phylogenetic tree of severe acute respiratory syndrome-related coronaviruses (SARSr-CoVs) from bats. SARSr-CoVs cluster into three lineages, L1–L3, and human severe acute respiratory syndrome coronaviruses (SARS-CoVs) embed in L1. Two individual SARSr-CoVs do not cluster into these lineages: YN, a virus isolated from Yunnan province, China, and BG, a virus from Bulgaria, Europe. The tree is based on published trees[20,138] and reconstructed using sequences of the complete RNA-dependent RNA polymerase-coding region (maximum likelihood method under the GTR+I+Γ model of nucleotide substitution as implemented in PhyML, version 3.1 (REF.[137])).The strain Zhejiang2013 (GenBank No. KF636752) was used as a root. **b** | By contrast, Middle East respiratory syndrome-related coronaviruses (MERSr-CoVs) form two major viral lineages, L1 and L2. L1 is found in humans and camels, and L2 is found only in camels. Two small clusters, B1 (bat 1) and B2, and one single virus, SA, from South Africa, were found in bats. The phylogenetic tree of MERSr-CoVs is based on a published trees[94,139] and reconstructed using full-genome alignment of all coding regions using the same method as above. HKU4-1 (EF065505) and HKU5-1 (EF065509), two 2c betacoronaviruses, served as the root of the tree. Detailed phylogenetic trees and grouping information can be found in Supplementary Fig. S1. MERS-CoV, Middle East respiratory syndrome coronavirus.
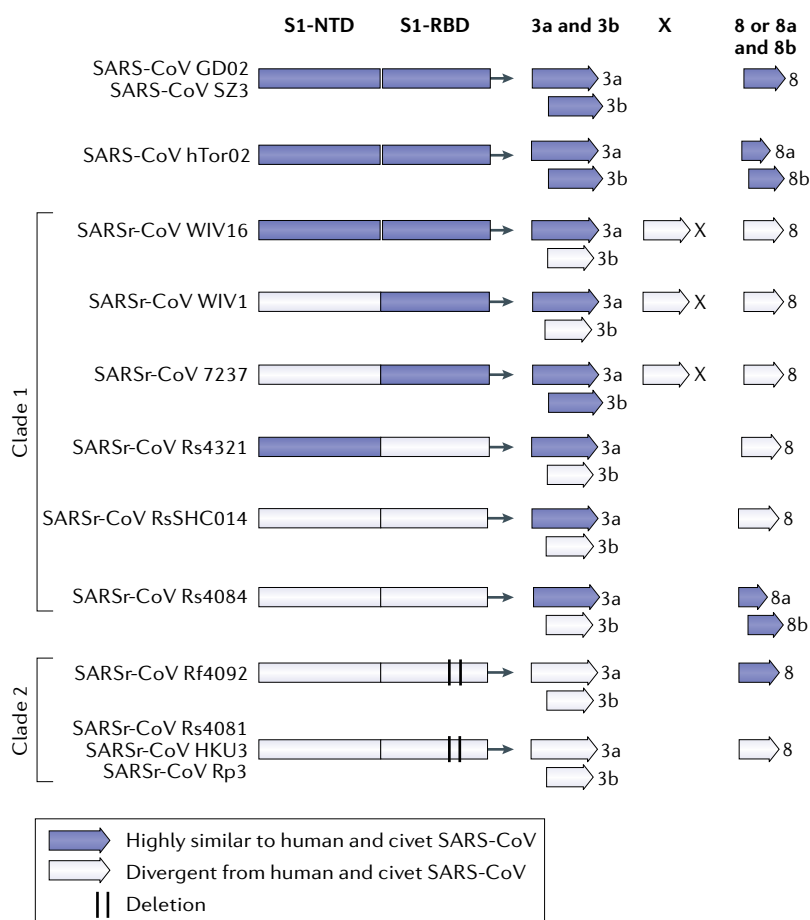
Fig. 5 | **Variable regions in different SARS-CoV and bat SARSr-CoV isolates.** Variability and thus species adaptation majorly affect three severe acute respiratory syndrome coronavirus (SARS-CoV) and SARS-related coronavirus (SARSr-CoV) proteins: the spike protein (S) (both the S1 amino-terminal domain (S1-NTD) and the S1 receptor-binding domain (S1-RBD) show variability), ORF3 (3a and 3b) and ORF8 (8a and b). SARS-CoV GD02 and hTor02 represent strains that were isolated from patients during the early, and middle or late phase of the SARS epidemic in 2002–2003, respectively; SARS-CoV CZ3 is a representative of strains isolated from civets in 2003 and 2004 (REFS[42,62]). All bat SARSr-CoVs, except HKU3 and Rp3, were discovered in Yunnan province during 2011–2015. On the basis of deletions in the RBD, bat SARSr-CoVs can be divided into two clades. Those without a deletion and thus an identical size in S1 to SARS-CoV can be further divided into four genotypes: genotype 1, represented by WIV16, is highly similar to SARS-CoV in both the NTD and the RBD; genotype 2, represented by WIV1, differs in NTD from SARS-CoV; genotype 3, represented by Rs4231, differs in RBD from SARS-CoV; and genotype 4, represented by SHC014 and Rs4084, differs in both NTD and RBD from SARS-CoV[20]. The differences in S influence species-specific receptor binding, whereas differences in the accessory proteins, including potentially the newly discovered ORFX (X), mainly affect immune responses and viral immune evasion. Adapted from REF.[20], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Salt bridge
A structure in proteins that forms a bond between oppositely charged residues that are sufficiently close to each other to experience electrostatic attraction.

motif (RBM)[82,141] (FIG. 6a). Two virus-binding hot spots have been identified at the interface of the RBD and human ACE2, centring on ACE2 residues Lys31 (hot spot 31) and Lys353 (hot spot 353)[83,84] (FIG. 6b). They both consist of a salt bridge (between Lys31 and Glu35 for hot spot 31 and between Lys353 and Asp38 for hot spot 353); both salt bridges are buried in hydrophobic pockets and contribute a substantial amount of energy to RBD–ACE2 binding as well as filling voids at the RBD–ACE2 interface. Naturally selected RBM mutations all interact with the hot spots (FIG. 6b; TABLE 1) and affect RBD–ACE2 binding.

Mutations in RBM residue 479 had an important role in the civet-to-human transmission of SARS-CoV[42,76,78,85]. Residue 479 is an asparagine in strains hTor02, hcGd03 and hHae08 but is a lysine in strain cSz02 and an arginine in strain cHb05 (TABLE 1). Asn479 is located near hot spot 31, without interfering with the structure of hot spot 31 (REF.[85]) (FIG. 6b, c). However, a change to Lys479 leads to steric and electrostatic interference with hot spot 31, reducing the binding affinity between the SARS-CoV RBD and human ACE2. By contrast, Arg479 reaches the vicinity of hot spot 353 and forms a salt bridge with ACE2 residue Asp38 (REF.[83]) (FIG. 6d). Hence, strains hTor02, hcGd03 and hHae08 (all of which contain Asn479) and strain cHb05 (which contains Arg479) recognize human ACE2 and infect human cells efficiently, whereas strain cSz02 (which contains Lys479) recognizes human ACE2 inefficiently and infects human cells inefficiently. The above structural analyses are supported by biochemical, functional and epidemiological data[42,76,78,83–85]. Because of residue differences between human ACE2 and civet ACE2, both Asn479 and Lys479 fit well into the interface between the RBD and civet ACE2, although Arg479 fits even better[83,85]; consequently, strains hTor02, cSz02, hcGd03 and cHb05 (which contain either Asn479, Lys479 or Arg479) recognize civet ACE2 and infect civet cells efficiently[79]. In sum, Asn479 and Arg479 are viral adaptations to human ACE2, whereas Lys479 is incompatible with human ACE2; Arg479 is a viral adaptation to civet ACE2, whereas Asn479 and Lys479 are also compatible with civet ACE2.

Mutations in RBM residue 487 had an important role in the human-to-human transmission of SARS-CoV. Residue 487 is a threonine in strain hTor02 but is a serine in the other strains isolated from humans and civets. The methyl group of Thr487 interacts with hot spot 353 in human ACE2 by providing stacking support for the formation of the salt bridge between Lys353 and Asp38; consequently, strain hTor02 recognizes human ACE2 efficiently and was transmitted between humans during the 2002–2003 SARS epidemic. By contrast, Ser487 cannot provide support to hot spot 353, and hence the other strains isolated from humans and civets recognize human ACE2 inefficiently. Consequently, neither cSz02 nor hcGd03 was transmitted between humans. The above structural analyses are supported by biochemical, functional and epidemiological data[42,76,78,83–85]. Because of residue differences between human ACE2 and civet ACE2, Ser487 fits well into the RBD–civet ACE2 interface although still not as well as Thr487 (REFS[83,85]); consequently, strains sSZ02, hcGd03 and cHb05 (which contain Ser487) recognize civet ACE2 and infect civet cells efficiently[79]. In sum, Thr487 is a viral adaptation to both human and civet ACE2, and Ser487 is much more compatible with civet ACE2 than with human ACE2 (FIG. 6b).

RBM residues 442, 472 and 480 also contribute to receptor recognition and host range of SARS-CoV although not as much as residues 479 and 487. Detailed structural, biochemical and functional analyses showed that Phe442, Phe472 and Asp480 are viral adaptations to human ACE2, whereas Tyr442, Leu472 or Pro472, and Gly480 are viral adaptations to civet ACE2 (REFS[72,83]).

To corroborate the importance of these residues for SARS-CoV binding to either human or civet ACE2, two SARS-CoV S proteins, hOptimize and cOptimize, were rationally designed: the former contains all of the human ACE2-adapted residues (Phe442, Phe472, Asn479, Asp480 and Thr487), whereas the latter contains the civet ACE2-adapted residues (Tyr442, Pro472, Arg479, Gly480 and Thr487). These two S proteins demonstrate exceptionally high affinity for human ACE2 and civet ACE2, confirming that the human ACE2-adapted and civet ACE2-adapted RBM residues help determine SARS-CoV host range[72,83]. In addition to
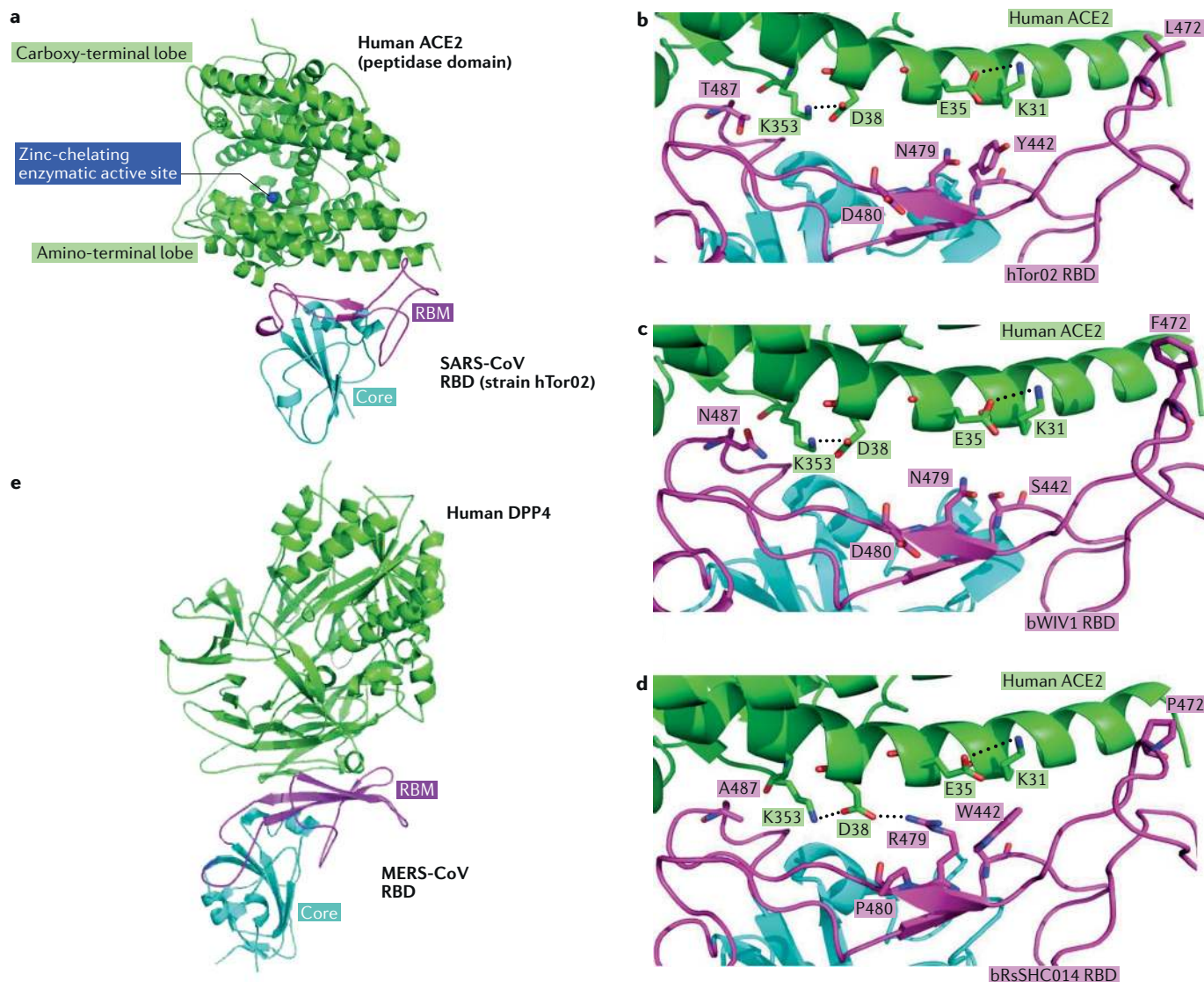
Fig. 6 | **Receptor recognition by SARS-CoV and MERS-CoV. a** | Severe acute respiratory syndrome coronavirus (SARS-CoV) uses its receptor-binding domain (RBD) (as shown in the structure of strain hTor02, containing core structure (cyan) and receptor-binding motif (RBM; magenta)) to bind human angiotensin-converting enzyme 2 (ACE2; green; Protein Data Bank ID: 2AJF). ACE2 is a peptidase with zinc (blue) in its active centre. **b** | Several residues in the host and viral receptor, as well as two salt bridges that stabilize the structure (dotted lines) and form two binding hot spots, are crucial for binding of the severe acute respiratory syndrome (SARS) epidemic strain hTor02. Hydrophobic residues surrounding the two salt bridges are present in the structure but are not shown in the figure. **c** | By contrast, the SARS-related coronavirus (SARSr-CoV) strain bWIV1, which was isolated from bats and can infect both civet and human cells, differs in residues 442, 472 and 487. The mutation from threonine to asparagine in residue 487 introduces a polar side chain and is predicted to interfere with binding at hot spot 353. The model shown here was built on the basis of the structure of hTor02 RBD complexed with human ACE2 (Protein Data

Bank ID: 2AJF), in which residues 442, 472 and 487 were mutated from those in strain hTor02 to those in strain bWIV1. **d** | The bat SARSr-CoV strain bRsSHC014 can also infect human and civet cells; it carries an alanine in position 487, and the short side chain of this residue does not support the structure of hot spot 353. The model was built on the basis of the structure of cOptimize RBD complexed with human ACE2 (Protein Data Bank ID: 3SCJ), in which residues 442, 480 and 487 were mutated from those in strain cOptimize to those in strain bWIV1. **e** | The Middle East respiratory syndrome coronavirus (MERS-CoV) RBD (core structure in cyan and RBM in magenta) binds human dipeptidyl peptidase 4 (DPP4; green; Protein Data Bank ID: 4KR0). Structure figures were made using PyMOL[115]. Modelled mutations in panels **c** and **d** were performed using Coot[140]. Panels **a–d** are adapted from REF.[83]: this research was originally published in *The Journal of Biological Chemistry*. Wu, K. L., Peng, G. Q., Wilken, M., Geraghty, R. J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J. Biol. Chem.* 2012; **287**:8904–8911. © American Society for Biochemistry and Molecular Biology.

Table 1 | **Mutations in the receptor-binding motif of SARS-CoV**

Strain	Host	Year	Truncation in receptor-binding motif	Residue 442	Residue 472	Residue 479	Residue 480	Residue 487
hTor02	Human	2002–2003	No	Tyr	Leu	Asn	Asp	Thr
cSz02	Civet	2002–2003	No	Tyr	Leu	Lys	Asp	Ser
hcGd03	Human and civet	2003–2004	No	Tyr	Pro	Asn	Gly	Ser
cHb05	Civet	2005	No	Tyr	Pro	Arg	Gly	Ser
hHae08	Human	2008	No	Phe	Phe	Asn	Asp	Ser
hOptimize	Human	In vitro design	No	Phe	Phe	Asn	Asp	Ser
cOptimize	Civet	In vitro design	No	Tyr	Pro	Arg	Gly	Thr
bHKU3	Bat	2005	Yes	Ser	Gly	Ser	Thr	Val
bWIV1	Bat	2013	No	Ser	Phe	Asn	Asp	Asn
bRsSHC014	Bat	2013	No	Trp	Pro	Arg	Pro	Ala

Data from Hu et al. 2012 (REF.[83]). SARS-CoV, severe acute respiratory syndrome coronavirus.

receptor binding, proteolytic cleavage of S and potentially other mutations that affect virion and trimer stability may also be important for virus transmissibility in different hosts, and these factors need to be studied further.

SARSr-CoV mutations that affect receptor binding. To date, numerous SARSr-CoV strains have been identified from bats[15,16,18–20]. These bat SARSr-CoVs are the likely progenitors of SARS-CoV that infected humans and civets, and hence understanding their interactions with human or civet ACE2 is critical for tracing the origins of SARS-CoV and for preventing and controlling future SARS-CoV outbreaks in humans. The RBD sequences of these bat SARSr-CoVs fall into three major groups; the representative strains from each group are bHKU3 (isolated in 2005), bWIV1 (isolated in 2013) and bRsSHC014 (isolated in 2013) (TABLE 1). Strains bWIV1 and bRsSHC014, but not strain bHKU3, use both human and civet ACE2 and hence can infect both human and civet cells[16,18–20,86,87]. Strain bHKU3 has a truncated RBM (TABLE 1), which distorts the structure of the RBM and abolishes its binding to human and civet ACE2. Neither strain bWIV1 nor strain bRsSHC014 contains truncations in its RBM, and hence, their RBMs likely retain the same structure as SARS-CoV RBMs. Here, we analysed the potential interactions between these two strains (bWIV1 and bRsSHC014) and human ACE2 by building homology structural models of their RBDs complexed with human ACE2, focusing on residues 479 and 487 (FIG. 6c, d). Strain bWIV1 contains Asn479 and Asn487 in its RBM. Whereas Asn479 is a viral adaptation to human ACE2, the polar side chain of Asn487 may have unfavourable interactions with the aliphatic portion of residue Lys353 in human ACE2, which is part of hot spot 353 (FIG. 6c). Strain bRsSHC014 contains Arg479 and Ala487 in its RBM. Whereas Arg479 is a viral adaptation to human ACE2, the small side chain of Ala487 does not provide support to the structure of hot spot 353 (FIG. 6d). Therefore, although both bWIV1 and bRsSHC014 can infect human airway cells, they bind human ACE2 less well than hTor02 and produce less severe symptoms than the epidemic strain of SARS-CoV in vivo[88,89]. Similarly, both bWIV1

and bRsSHC014 can infect civet cells, but they bind civet ACE2 less well than cSz02. Thus, it is predicted that both strains will be attenuated compared with early-phase or late-phase human SARS epidemic viruses. Future evolution of bat SARSr-CoV strains bWIV1 and bRsSHC014 in crucial RBM residues may allow them to cross the species barriers between bats, civets and humans, posing potential health threats.

Origin and evolution of MERS-CoV

Whereas the emergence of SARS involved palm civets, most of the early MERS index cases had contact with dromedary camels. Indeed, MERS-CoV strains isolated from camels were almost identical to those isolated from humans[90–95]. Moreover, MERS-CoV-specific antibodies were highly prevalent in camels from the Middle East, Africa and Asia[13,14,96–103]. MERS-CoV infections were detected in camel serum samples collected in 1983 (REF.[100]), suggesting that MERS-CoV was present in camels at least 30 years ago. Genomic sequence analysis indicated that MERS-CoV, *Tylonycteris* bat coronavirus HKU4 and *Pipistrellus* bat coronavirus HKU5 are phylogenetically related (denoted as betacoronavirus lineage C)[21]. The viruses in this lineage have identical genomic structures and are highly conserved in their polyproteins and most structural proteins, but their S proteins and accessory proteins are highly variable. MERSr-CoVs were found in at least 14 bat species from two bat families, Vespertilionidae and Nycteridae. However, none of these MERSr-CoVs is a direct progenitor of MERS-CoV, as their S proteins differ substantially from that of MERS-CoV[98,104–106].

To understand the evolutionary relationships between MERS-CoV and MERSr-CoVs, we constructed a phylogenetic tree on the basis of the alignment of all the coding regions (FIG. 4b; Supplementary Fig. S1b). The phylogenetic tree contains two main clusters and several small clades or strains. Overall, the genetic diversity within the L1 and L2 viral lineages is low, indicating that humans and camels have been infected by viruses from the same source within a short time period. The L1 viruses include human and camel MERS-CoVs mainly

from the Middle East (the United Arab Emirates, the Kingdom of Saudi Arabia, Oman and Jordan) and two Asian countries (South Korea and Thailand) that had caused outbreaks in human populations. It is worth noting that the cases reported in South Korea and Thailand were related to those in the Middle East. The L2 viruses include camel MERS-CoVs from Africa (Nigeria, Burkina Faso and Ethiopia) and one Middle East country (Morocco); these viruses have not caused any human infection. Clearly, these two viral lineages share a common ancestor but have diverged in their potential to cause human infections. The MERSr-CoV strain Neoromicia/5038 (GenBank No. MF593268) isolated in South Africa was the closest relative to MERS-CoVs in the phylogenetic tree. Overall, all the MERSr-CoVs isolated from bats support the hypothesis that MERS-CoV originated from bats. However, given the phylogenetic gap between the bat MERSr-CoVs and human and camel MERS-CoVs, there should be other yet-to-be-identified viruses that are circulating in nature and directly contributed to the emergence of MERS-CoV in humans and camels. Hopefully, such viruses will be found in bats in the future.

Not surprisingly, recombination events have taken place in the evolution and emergence of MERS-CoV[94,105,107–109]. Phylogenetic trees constructed using genes encoding *orf1ab* and *S* were incongruent with the tree topology of the complete genome, suggesting potential recombination in these genes[108]. Numerous recombinations imply that MERS-CoV originated from the exchange of genetic elements between different viral ancestors, including those isolated from camels and the assumed natural host bats[94,105,107,110,111].

Variability of human and camel MERS-CoV
The full-length genomic sequences of MERS-CoVs isolated from humans and camels are almost identical (>99% identity). The major variations are located in S, ORF4b and ORF3, particularly in African camel MERS-CoVs[94]. Substitutions of a few amino acid residues were found in the S protein of some camel MERS-CoVs, but none of them was located in the RBD[94,112]. Neutralization assays indicated that camel sera that are positive for MERS-CoV can completely neutralize the human MERS-CoV strains, suggesting that MERS-CoVs isolated from humans and camels are antigenically similar to each other[94]. MERS-CoVs from both humans and camels contain variable ORF3 and ORF4 proteins with different lengths owing to either terminal truncations or internal deletions[94]. ORF4b is known to be an interferon antagonist[113,114]. MERS-CoV isolates from West African camels with a truncated ORF4b gene replicate less efficiently in human cell culture and are less pathogenic in human DPP4 transgenic mice[94]. Curiously, deletion of the *orf4* gene in the human MERS-CoV strain EMC did not substantially reduce virus replication, although it induced a stronger interferon response[94]. Another study demonstrated that the deletion of *orf3–orf5* dramatically attenuated MERS-CoV virulence, primarily through increased host responses, including disrupted cellular processes, increased activation of the interferon pathway and robust inflammation[115].

Variability of bat MERSr-CoVs
To date, bat MERSr-CoVs and human and camel MERS-CoVs share the same genomic structures but differ substantially in their genomic sequences[105,106,110,111,116]. The highest overall genomic sequence identity between bat MERSr-CoV and human and camel MERS-CoV is ~85%. On the basis of their genomic sequences, several bat MERSr-CoV strains discovered in China, such as Ii-MERSr-CoV, Ve-MERSr-CoV and Hy-MERSr-CoV, have just reached the taxonomic threshold to be considered the same species as MERS-CoV[106,110,111].

Compared with human and camel MERS-CoV, bat MERSr-CoVs vary most in S and accessory genes. The sequence identity of the S protein between bat MERSr-CoVs and human and camel MERS-CoVs is approximately 45–65%, with even lower sequence identity in the RBD region[110,111]. The size of these S proteins differs in these strains, mainly because of deletions in their RBD region and/or the S1 and S2 boundary. These deletions are considered to be related to the differences in receptor binding and cell entry[111,116]. The accessory genes, including those encoding ORF3, ORF4a, ORF4b and ORF5, are also highly variable in length and sequence between bat MERSr-CoVs and human and camel MERS-CoVs, suggesting substantial evolution of these genes in their natural hosts[105,106,110,111,116].

Receptor usage of MERS-CoV and MERSr-CoV
In contrast to SARS-CoV, which uses ACE2 as its receptor, MERS-CoV uses DPP4. Similar to SARS-CoV S1-CTD, MERS-CoV S1-CTD functions as the viral RBD[10,117]. Like the SARS-CoV S1-CTD, the MERS-CoV S1-CTD also contains two subdomains, a core structure and an RBM[9,118–120] (FIG. 6e). The core structures of these two S1-CTDs are similar to each other, with both containing a five-stranded β-sheet as the main scaffold. However, their RBMs differ substantially: whereas the SARS-CoV RBM mainly contains loops, the MERS-CoV RBM mainly contains a four-stranded β-sheet. The structural differences between MERS-CoV and SARS-CoV RBMs account for the different receptor specificities of the two viruses[121].

Like the interactions between SARS-CoV and ACE2, the interactions between MERS-CoV and DPP4 have been extensively examined. DPP4 from humans, camels, horses and bats can function as a receptor for MERS-CoV, whereas DPP4 from mice, hamsters and ferrets cannot[112,122–125]. Key residue differences between human DPP4 and the DPP4 from other species affect the species specificities of MERS-CoV. For example, two residues (288 and 330) in mouse DPP4 and five residues (291, 295, 336, 341 and 346) in hamster DPP4 are largely responsible for the incompatibility of mouse and hamster DPP4s with MERS-CoV[112,123]. Mutating these residues to the corresponding residues in human DPP4 makes mouse and hamster DPP4 functional receptors for MERS-CoV. On the other hand, MERS-CoV and MERSr-CoVs have been isolated from camels and bats, respectively. MERS-CoV strains isolated from humans and camels are highly similar to each other, and they both use human DPP4 efficiently[112]. MERSr-CoVs from bats in general share only ~60–70% sequence identity with MERS-CoV in the RBD, and only some of these bat viruses, including HKU4,

recognize DPP4 as the receptor[110,111,126]. However, they bind DPP4 less efficiently than MERS-CoV. Mutating three residues in the HKU4 RBD (540, 547 and 558) substantially increased its affinity for human DPP4 (REF.[127]). Overall, as in the case of SARS-CoV, receptor recognition is a crucial determinant of the host range of MERS-CoV.

SADS-CoV
From 28 October 2016 to 2 May 2017, swine acute diarrhoea syndrome (SADS) was observed in four pig breeding farms in Guangdong province, with a mortality up to 90% for piglets 5 days or younger. A novel HKU2-related bat coronavirus, named SADS-CoV, was identified as the causative agent[34,128,129]. The SADS-CoV isolates from piglets of the four farms were almost identical and shared 95% identity with *Rhinolophus* bat coronavirus HKU2 (REF.[130]), indicating the bat origin of this pig virus. Immediately after the SADS outbreak, SADS-related CoVs (SADSr-CoVs) with 96–98% sequence identity to SADS-CoV were detected in 9.8% of anal swabs collected from different *Rhinolophus* species in Guangdong province during 2013–2016. Although genetically highly similar, bat SADSr-CoVs show high genetic diversity in the *S* gene, with 72–92% nucleotide and 80–98% amino acid identity to SADS-CoV. Receptor analysis indicated that none of the known coronavirus receptors, ACE2, DPP4 and aminopeptidase N, are essential for SADS-CoV entry[34]. The mechanism of transmission of SADS-CoV from bats to pigs and the pathogenesis of bat-originated SADSr-CoVs in pigs need further exploration. This is the first documented spillover of a bat coronavirus that caused severe diseases in domestic animals, although molecular evolution data suggested PEDV probably originated in bats[37,38].

Conclusions and future perspectives
The collected data on genetic evolution, receptor binding and pathogenesis demonstrated that SARS-CoV most likely originated in bats through sequential recombination of bat SARSr-CoVs. Recombination likely occurred in bats before SARS-CoV was introduced into Guangdong province through infected civets or other infected mammals from Yunnan. The introduced SARS-CoV underwent rapid mutations in *S* and *orf8* and successfully spread in market civets. After several independent spillovers to humans, some of the strains underwent further mutations in *S* and became epidemic during the SARS outbreak in 2002–2003. However, a recent serological investigation revealed the presence of antibodies against the SARSr-CoV nucleocapsid in humans living around a bat cave but who had not shown clinical signs of disease, suggesting that the virus can infect humans through frequent contact[131].

A similar scenario might have happened for MERS-CoV. Since its outbreak in 2012, MERSr-CoVs and related viruses (HKU4 and HKU5) have been found in different bat species in five continents[17,21,106,110,111,116,126,127,132]. The ORF1ab of these viruses is highly similar to MERS-CoV ORF1ab, but they are highly diverse in their S proteins. Surprisingly, some bat MERSr-CoVs and HKU can use the same receptor, DPP4, as MERS-CoV[110,111,126,127]. Given the massive number of coronaviruses carried by different bat species, the high plasticity in receptor usage and other features such as adaptive mutation and recombination, frequent interspecies transmission from bats to animals and humans is expected.

Currently, no clinical treatments or prevention strategies are available for any human coronavirus. Given the conserved RBDs of SARS-CoV and bat SARSr-CoVs, some anti-SARS-CoV strategies in development, such as anti-RBD antibodies or RBD-based vaccines, should be tested against bat SARSr-CoVs. Recent studies demonstrated that anti-SARS-CoV strategies worked against only WIV1 and not SHC014 (REFS[71,88,89]). In addition, little information is available on HKU3-related strains that have much wider geographical distribution and bear truncations in their RBD. Similarly, anti-S antibodies against MERS-CoV could not protect from infection with a pseudovirus bearing the bat MERSr-CoV S[111]. Furthermore, little is known about the replication and pathogenesis of these bat viruses. Thus, future work should be focused on the biological properties of these viruses using virus isolation, reverse genetics and in vitro and in vivo infection assays. The resulting data would help the prevention and control of emerging SARS-like or MERS-like diseases in the future.

It is widely accepted that many viruses have existed in their natural reservoirs for a very long time. The constant spillover of viruses from natural hosts to humans and other animals is largely due to human activities, including modern agricultural practices and urbanization. Therefore, the most effective way to prevent viral zoonosis is to maintain the barriers between natural reservoirs and human society, in mind of the 'one health' concept.

1. Masters, P. S. & Perlman, S. in *Fields Virology* Vol. 2 (eds Knipe, D. M. & Howley, P. M.) 825–858 (Lippincott Williams & Wilkins, 2013).
2. Zhong, N. S. et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**, 1353–1358 (2003).
3. Drosten, C. et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003).
4. Fouchier, R. A. et al. Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* **423**, 240 (2003).
5. Ksiazek, T. G. et al. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953–1966 (2003).
6. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
7. Li, W. H. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
8. Qian, Z. et al. Innate immune response of human alveolar type II cells infected with severe acute respiratory syndrome-coronavirus. *Am. J. Respir. Cell. Mol. Biol.* **48**, 742–748 (2013).
9. Lu, G. et al. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* **500**, 227–231 (2013).
10. Raj, V. S. et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495**, 251–254 (2013).
11. Scobey, T. et al. Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. *Proc. Natl Acad. Sci. USA* **110**, 16157–16162 (2013).
12. Guan, Y. et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276–278 (2003). **This paper is the first demonstration of SARS-CoV transmission from animals.**
13. Alagaili, A. N. et al. Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. *MBio* **5**, e00884–14 (2014).
14. Hemida, M. G. et al. Middle East Respiratory Syndrome (MERS) coronavirus seroprevalence in domestic livestock in Saudi Arabia, 2010 to 2013. *Euro. Surveill.* **18**, 21–27 (2013).
15. Lau, S. K. et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl Acad. Sci. USA* **102**, 14040–14045 (2005).

16. Kan, B. et al. 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 (2005).

17. Ithete, N. L. et al. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerg. Infect. Dis.* **19**, 1697–1699 (2013).

18. Ge, X. Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).
This paper provides the first identification of human ACE2 as the receptor for bat SARS-like coronavirus.

19. Yang, X. L. et al. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. *J. Virol.* **90**, 3253–3256 (2016).

20. Hu, B. et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLOS Pathog.* **13**, e1006698 (2017).
This paper identifies a gene pool of SARS-CoVs in bats.

21. Lau, S. K. et al. Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. *J. Virol.* **87**, 8638–8650 (2013).

22. Woo, P. C., Lau, S. K., Huang, Y. & Yuen, K. Y. Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med. (Maywood)* **234**, 1117–1127 (2009).

23. Perlman, S. & Netland, J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.* **7**, 439–450 (2009).

24. Woo, P. C. et al. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J. Virol.* **86**, 3995–4008 (2012).
This paper describes coronavirus origins by phylogenetic analysis.

25. Graham, R. L., Donaldson, E. F. & Baric, R. S. A decade after SARS: strategies for controlling emerging coronaviruses. *Nat. Rev. Microbiol.* **11**, 836–848 (2013).

26. Hu, B., Ge, X., Wang, L. F. & Shi, Z. Bat origin of human coronaviruses. *Virol. J.* **12**, 221 (2015).

27. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).

28. Su, S. et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* **24**, 490–502 (2016).

29. Forni, D., Cagliani, R., Clerici, M. & Sironi, M. Molecular evolution of human coronavirus genomes. *Trends Microbiol.* **25**, 35–48 (2017).

30. Anthony, S. J. et al. Global patterns in coronavirus diversity. *Virus Evol.* **3**, vex012 (2017).

31. Wang, L., Su, S., Bi, Y., Wong, G. & Gao, G. F. Bat-origin coronaviruses expand their host range to pigs. *Trends Microbiol.* **26**, 466–470 (2018).

32. Brian, D. A. & Baric, R. S. Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* **287**, 1–30 (2005).

33. Lin, C. M., Saif, L. J., Marthaler, D. & Wang, Q. Evolution, antigenicity and pathogenicity of global porcine epidemic diarrhea virus strains. *Virus Res.* **226**, 20–39 (2016).

34. Zhou, P. et al. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* **556**, 255–258 (2018).
This paper describes bat origin of a swine coronavirus.

35. Huang, Y. W. et al. Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States. *MBio* **4**, e00737–00713 (2013).

36. Liu, C. et al. Receptor usage and cell entry of porcine epidemic diarrhea coronavirus. *J. Virol.* **89**, 6121–6125 (2015).

37. Simas, P. V. et al. Bat coronavirus in Brazil related to appalachian ridge and porcine epidemic diarrhea viruses. *Emerg. Infect. Dis.* **21**, 729–731 (2015).

38. Lacroix, A. et al. Genetic diversity of coronaviruses in bats in Lao PDR and Cambodia. *Infect. Genet. Evol.* **48**, 10–18 (2017).

39. Tu, C. et al. Antibodies to SARS coronavirus in civets. *Emerg. Infect. Dis.* **10**, 2244–2248 (2004).

40. Wang, M. et al. [Analysis on the risk factors of severe acute respiratory syndromes coronavirus infection in workers from animal markets]. *Zhonghua Liu Xing Bing Xue Za Zhi* **25**, 503–505 (2004).

41. Xu, H. F. et al. [An epidemiologic investigation on infection with severe acute respiratory syndrome coronavirus in wild animals traders in Guangzhou]. *Zhonghua Yu Fang Yi Xue Za Zhi* **38**, 81–83 (2004).

42. Song, H. D. et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl Acad. Sci. USA* **102**, 2430–2435 (2005).
This paper describes genetic evolution of SARS-CoV during transmission from animals and humans.

43. Li, W. et al. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).

44. Ren, W. et al. Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. *J. Gen. Virol.* **87**, 3355–3359 (2006).

45. Drexler, J. F. et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J. Virol.* **84**, 11336–11349 (2010).

46. Lau, S. K. et al. 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 (2010).

47. Rihtaric, D., Hostnik, P., Steyer, A., Grom, J. & Toplak, I. Identification of SARS-like coronaviruses in horseshoe bats (Rhinolophus hipposideros) in Slovenia. *Arch. Virol.* **155**, 507–514 (2010).

48. Yuan, J. et al. Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans. *J. Gen. Virol.* **91**, 1058–1062 (2010).

49. Balboni, A., Gallina, L., Palladini, A., Prosperi, S. & Battilani, M. A real-time PCR assay for bat SARS-like coronavirus detection and its application to Italian greater horseshoe bat faecal sample surveys. *Sci. World J.* **2012**, 989514 (2012).

50. Yang, L. et al. Novel SARS-like betacoronaviruses in bats, China, 2011. *Emerg. Infect. Dis.* **19**, 989–991 (2013).

51. He, B. et al. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. *J. Virol.* **88**, 7070–7082 (2014).

52. Gouilh, M. A. et al. SARS-coronavirus ancestor's foot-prints in South-East Asian bat colonies and the refuge theory. *Infect. Genet. Evol.* **11**, 1690–1702 (2011).

53. Wacharapluesadee, S. et al. Diversity of coronavirus in bats from Eastern Thailand. *Virol. J.* **12**, 57 (2015).

54. Tong, S. et al. Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg. Infect. Dis.* **15**, 482–485 (2009).

55. Wang, M. N. et al. Longitudinal surveillance of SARS-like coronaviruses in bats by quantitative real-time PCR. *Virol. Sin.* **31**, 78–80 (2016).

56. Lai, M. M. & Cavanagh, D. The molecular biology of coronaviruses. *Adv. Virus Res.* **48**, 1–100 (1997).

57. Zhao, Z. et al. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol. Biol.* **4**, 21 (2004).

58. Hon, C. et al. Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J. Virol.* **82**, 1819–1826 (2008).

59. Wu, Z. et al. ORF8-related genetic evidence for chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J. Infect. Dis.* **213**, 579–583 (2016).

60. Nagy, P. D. & Simon, A. E. New insights into the mechanisms of RNA recombination. *Virology* **235**, 1–9 (1997).

61. Rowe, C. L. et al. Generation of coronavirus spike deletion variants by high-frequency recombination at regions of predicted RNA secondary structure. *J. Virol.* **71**, 6183–6190 (1997).

62. Chinese, S. M. E. C. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **303**, 1666–1669 (2004).
This paper describes the genetic evolution of human SARS-CoV during SARS outbreaks.

63. Babcock, G. J., Esshaki, D. J., Thomas, W. D. & Ambrosino, D. M. Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor. *J. Virol.* **78**, 4552–4560 (2004).

64. Wong, S. K., Li, W. H., Moore, M. J., Choe, H. & Farzan, M. A. 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J. Biol. Chem.* **279**, 3197–3201 (2004).
This paper identifies the receptor-binding domain of the SARS-CoV spike protein.

65. Le, T. M. et al. Expression, post-translational modification and biochemical characterization of proteins encoded by subgenomic mRNA8 of the severe acute respiratory syndrome coronavirus. *FEBS J.* **274**, 4211–4222 (2007).

66. Oostra, M., de Haan, C. A. & Rottier, P. J. The 29-nucleotide deletion present in human but not in animal severe acute respiratory syndrome coronaviruses disrupts the functional expression of open reading frame 8. *J. Virol.* **81**, 13876–13888 (2007).

67. Wong, H. H. et al. Accessory proteins 8b and 8ab of severe acute respiratory syndrome coronavirus suppress the interferon signaling pathway by mediating ubiquitin-dependent rapid degradation of interferon regulatory factor 3. *Virology* **515**, 165–175 (2018).

68. Sung, S. C., Chao, C. Y., Jeng, K. S., Yang, J. Y. & Lai, M. M. The 8ab protein of SARS-CoV is a luminal ER membrane-associated protein and induces the activation of ATF6. *Virology* **387**, 402–413 (2009).

69. Chen, C. Y. et al. Open reading frame 8a of the human severe acute respiratory syndrome coronavirus not only promotes viral replication but also induces apoptosis. *J. Infect. Dis.* **196**, 405–415 (2007).

70. Zhou, P., Li, H., Wang, H., Wang, L. F. & Shi, Z. Bat severe acute respiratory syndrome-like coronavirus ORF3b homologues display different interferon antagonist activities. *J. Gen. Virol.* **93**, 275–281 (2012).

71. Zeng, L. P. et al. Cross-neutralization of SARS coronavirus-specific antibodies against bat SARS-like coronaviruses. *Sci. China Life Sci.* **60**, 1399–1402 (2017).

72. Li, F. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Res.* **100**, 246–254 (2013).

73. Li, W. H. et al. Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions. *J. Virol.* **80**, 4211–4219 (2006).

74. Li, F. et al. Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *J. Virol.* **80**, 6794–6800 (2006).

75. Yuan, Y. et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **8**, 15092 (2017).

76. Li, W. H. et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* **24**, 1634–1643 (2005).
This paper identifies key residues involved in SARS-CoV adaptation to humans.

77. Towler, P. et al. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *J. Biol. Chem.* **279**, 17996–18007 (2004).

78. Qu, X. X. et al. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *J. Biol. Chem.* **280**, 29588–29595 (2005).

79. Liu, L. et al. Natural mutations in the receptor binding domain of spike glycoprotein determine the reactivity of cross-neutralization between palm civet coronavirus and severe acute respiratory syndrome coronavirus. *J. Virol.* **81**, 4694–4700 (2007).

80. Liang, G. D. et al. Laboratory diagnosis of four recent sporadic cases of community-acquired SARS, guangdong province, china. *Emerg. Infect. Dis.* **10**, 1774–1781 (2004).

81. Sheahan, T. et al. Mechanisms of zoonotic severe acute respiratory syndrome coronavirus host range expansion in human airway epithelium. *J. Virol.* **82**, 2274–2285 (2008).

82. Li, F., Li, W. H., Farzan, M. & Harrison, S. C. in *The Nidoviruses: Toward Control of SARSs and Other Nidovirus Diseases* (Advances in Experimental Medicine and Biology Vol. 581), 229–234 (Springer, 2006).

83. Wu, K. L., Peng, G. Q., Wilken, M., Geraghty, R. J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J. Bio. Chem.* **287**, 8904–8911 (2012).

84. Wu, K. et al. A virus-binding hot spot on human angiotensin-converting enzyme 2 Is critical for binding of two different coronaviruses. *J. Virol.* **85**, 5331–5337 (2011).
85. Li, F. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *J. Virol.* **82**, 6984–6991 (2008).
86. Becker, M. M. et al. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl Acad. Sci. USA* **105**, 19944–19949 (2008).
87. Ren, W. et al. Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *J. Virol.* **82**, 1899–1907 (2008).
88. Menachery, V. D. et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **21**, 1508–1513 (2015).
89. Menachery, V. D. et al. SARS-like WIV1-CoV poised for human emergence. *Proc. Natl Acad. Sci. USA* **113**, 3048–3053 (2016).
90. Haagmans, B. L. et al. Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *Lancet Infect. Dis.* **14**, 140–145 (2014). **This paper provides the first identification of MERS-CoV in camels.**
91. Azhar, E. I. et al. Evidence for camel-to-human transmission of MERS coronavirus. *N. Engl. J. Med.* **370**, 2499–2505 (2014).
92. Raj, V. S. et al. Isolation of MERS coronavirus from a dromedary camel, Qatar, 2014. *Emerg. Infect. Dis.* **20**, 1339–1342 (2014).
93. Sabir, J. S. et al. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science* **351**, 81–84 (2016).
94. Chu, D. K. W. et al. MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. *Proc. Natl Acad. Sci. USA* **115**, 3144–3149 (2018). **This paper describes the genetic diversity of MERS-CoV in African camels.**
95. Paden, C. R. et al. Zoonotic origin and transmission of Middle East respiratory syndrome coronavirus in the UAE. *Zoo. Pub. Health* **65**, 322–333 (2018).
96. Perera, R. A. et al. Seroepidemiology for MERS coronavirus using microneutralisation and pseudoparticle virus neutralisation assays reveal a high prevalence of antibody in dromedary camels in Egypt, June 2013. *Euro. Surveill.* **18**, 20574 (2013).
97. Reusken, C. B. et al. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *Lancet Infect. Dis.* **13**, 859–866 (2013).
98. Hemida, M. G. et al. MERS coronavirus in dromedary camel herd, Saudi Arabia. *Emerg. Infect. Dis.* **20**, 1231–1234 (2014).
99. Corman, V. M. et al. Antibodies against MERS coronavirus in dromedary camels, Kenya, 1992–2013. *Emerg. Infect. Dis.* **20**, 1319–1322 (2014).
100. Muller, M. A. et al. MERS coronavirus neutralizing antibodies in camels, Eastern Africa, 1983–1997. *Emerg. Infect. Dis.* **20**, 2093–2095 (2014).
101. Muller, M. A. et al. Presence of Middle East respiratory syndrome coronavirus antibodies in Saudi Arabia: a nationwide, cross-sectional, serological study. *Lancet Infect. Dis.* **15**, 559–564 (2015).
102. Saqib, M. et al. Serologic evidence for MERS-CoV infection in dromedary camels, Punjab, Pakistan, 2012–2015. *Emerg. Infect. Dis.* **23**, 550–551 (2017).
103. Harcourt, J. L. et al. The prevalence of Middle East respiratory syndrome coronavirus (MERS-CoV) antibodies in dromedary camels in Israel. *Zoonoses Public Health* **65**, 749–754 (2018).
104. de Groot, R. J. et al. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J. Virol.* **87**, 7790–7792 (2013).
105. Corman, V. M. et al. Rooting the phylogenetic tree of middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. *J. Virol.* **88**, 11297–11303 (2014).
106. Yang, L. et al. MERS-related betacoronavirus in Vespertilio superans bats. *China. Emerg. Infect. Dis.* **20**, 1260–1262 (2014).

107. Dudas, G. & Rambaut, A. MERS-CoV recombination: implications about the reservoir and potential for adaptation. *Virus Evol.* **2**, vev023 (2016).
108. Wang, Y. et al. 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**, e01280–15 (2015).
109. Zhang, Z., Shen, L. & Gu, X. Evolutionary dynamics of MERS-CoV: potential recombination, positive selection and transmission. *Sci. Rep.* **6**, 25049 (2016).
110. Lau, S. K. P. et al. Receptor usage of a novel bat lineage C betacoronavirus reveals evolution of Middle East respiratory syndrome-related coronavirus spike proteins for human dipeptidyl peptidase 4 binding. *J. Infect. Dis.* **218**, 197–207 (2018).
111. Luo, C. M. et al. Discovery of novel bat coronaviruses in South China that use the same receptor as Middle East respiratory syndrome coronavirus. *J. Virol.* **92**, e00116–18 (2018).
112. Barlan, A. et al. Receptor variation and susceptibility to Middle East respiratory syndrome coronavirus infection. *J. Virol.* **88**, 4953–4961 (2014).
113. Matthews, K. L., Coleman, C. M., van der Meer, Y., Snijder, E. J. & Frieman, M. B. The ORF4b-encoded accessory proteins of Middle East respiratory syndrome coronavirus and two related bat coronaviruses localize to the nucleus and inhibit innate immune signalling. *J. Gen. Virol.* **95**, 874–882 (2014).
114. Yang, Y. et al. Middle East respiratory syndrome coronavirus ORF4b protein inhibits type I interferon production through both cytoplasmic and nuclear targets. *Sci. Rep.* **5**, 17554 (2015).
115. Menachery, V. D. et al. MERS-CoV accessory ORFs play key role for infection and pathogenesis. *MBio* **8**, e00665–17 (2017).
116. Anthony, S. J. et al. Further evidence for bats as the evolutionary source of Middle East respiratory syndrome coronavirus. *MBio* **8**, e00373–17 (2017).
117. Du, L. et al. Identification of a receptor-binding domain in the S protein of the novel human coronavirus Middle East respiratory syndrome coronavirus as an essential target for vaccine development. *J. Virol.* **87**, 9939–9942 (2013).
118. Mou, H. et al. The receptor binding domain of the new Middle East respiratory syndrome coronavirus maps to a 231-residue region in the spike protein that efficiently elicits neutralizing antibodies. *J. Virol.* **87**, 9379–9383 (2013).
119. Wang, N. et al. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res.* **23**, 986–993 (2013).
120. Chen, Y. et al. Crystal structure of the receptor-binding domain from newly emerged Middle East respiratory syndrome coronavirus. *J. Virol.* **87**, 10777–10783 (2013).
121. Li, F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J. Virol.* **89**, 1954–1964 (2015).
122. Cockrell, A. S. et al. Mouse dipeptidyl peptidase 4 is not a functional receptor for Middle East respiratory syndrome coronavirus infection. *J. Virol.* **88**, 5195–5199 (2014).
123. van Doremalen, N. et al. Host species restriction of Middle East respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4. *J. Virol.* **88**, 9220–9232 (2014).
124. Peck, K. M. et al. Glycosylation of mouse DPP4 plays a role in inhibiting Middle East respiratory syndrome coronavirus infection. *J. Virol.* **89**, 4696–4699 (2015).
125. Peck, K. M. et al. Permissivity of dipeptidyl peptidase 4 orthologs to Middle East respiratory syndrome coronavirus is governed by glycosylation and other complex determinants. *J. Virol.* **91**, e00534–17 (2017).
126. Yang, Y. et al. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. *Proc. Natl Acad. Sci. USA* **111**, 12516–12521 (2014).
127. Wang, Q. et al. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* **16**, 328–337 (2014).

128. Gong, L. et al. A new bat-HKU2-like coronavirus in swine, China, 2017. *Emerg. Infect. Dis.* **23**, 9 (2017).
129. Pan, Y. et al. Discovery of a novel swine enteric alphacoronavirus (SeACoV) in southern China. *Vet. Microbiol.* **211**, 15–21 (2017).
130. Lau, S. K. et al. Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. *Virology* **367**, 428–439 (2007).
131. Wang, N. et al. Serological evidence of bat SARS-related coronavirus infection in humans. *China. Virol. Sin.* **33**, 104–107 (2018).
132. Memish, Z. A. et al. Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg. Infect. Dis.* **19**, 1819–1823 (2013).
133. Huynh, J. et al. Evidence supporting a zoonotic origin of human coronavirus strain NL63. *J. Virol.* **86**, 12816–12825 (2012).
134. Tao, Y. et al. Surveillance of bat coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history. *J. Virol.* **91**, e01953–16 (2017). **This paper describes the bat origins of two human coronaviruses, NL63 and 229E.**
135. Corman, V. M. et al. Link of a ubiquitous human coronavirus to dromedary camels. *Proc. Natl Acad. Sci. USA* **113**, 9864–9869 (2016).
136. Drexler, J. F., Corman, V. M. & Drosten, C. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. *Antiviral Res.* **101**, 45–56 (2014).
137. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
138. Lin, X. D. et al. Extensive diversity of coronaviruses in bats from China. *Virology* **507**, 1–10 (2017).
139. Mackay, I. M. & Arden, K. E. MERS coronavirus: diagnostics, epidemiology and transmission. *Virol. J.* **12**, 222 (2015).
140. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
141. Li, F. et al. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).

Author contributions
All authors researched data for the article, contributed substantially to discussion of the content, wrote the article and reviewed and edited the manuscript before submission.

Competing interests
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Microbiology thanks R. Baric, B. Haagmans and K.-Y. Yuen for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41579-018-0118-9.

RELATED LINKS
International Committee on Taxonomy of Viruses: http://www.ictvonline.org/