Investigation on the human coronaviruses origin (bats and pangolins): a review

James Blackar MAWOLO* ☑, Caselia AKITI ☑, Harris Kollie MOMO ☑
Department of Biological Sciences, Thomas T.J.R. Faulkner College of Science and Technology, University of Liberia, Monrovia, Republic of Liberia

Received: 10.04.2020 • Accepted/Published Online: 31.08.2020 • Final Version: 14.09.2020

Abstract: A coronavirus related to SARS-CoV-2 has been isolated from Malayan pangolins illegally imported into Guangdong Province. It is not the precursor of SARS-CoV-2, but a comparison of viral genome sequences provides further evidence that the virus currently infecting humans. Bats and pangolins have been suggested as the natural reservoirs of a large variety of viruses. Some researchers have given attention to other species as the origin of coronaviruses and none have referred to bats and pangolins as the two emerging coronaviruses origin, which have caused unexpected human disease outbreaks recently. Severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV), are suggested to be originated from bats and pangolins. Numerous species of bats and pangolins in China have been observed to harbor genetically diverse SARS-like coronaviruses. Some strains are highly similar to SARS-CoV even in the spike protein and are able to use the same receptor as SARS-CoV for cell entry. Meanwhile, different coronaviruses phylogenetically related to MERS-CoV have been observed in the bats and pangolins species, some might be classified as similar to coronavirus species as MERS-CoV. Coronaviruses genetically related to human coronavirus 229E and NL63 have been found in bats and pangolins, respectively. However, intermediate hosts are suggested to play an important role in the transmission and emergence of these coronaviruses from bats and pangolins to humans. This study further documented that bats and pangolins origin of human coronaviruses are meaningful to predict and prevent a future outbreak of the deadly pandemic.

Key words: Bats, pangolins, coronavirus, infectious diseases, SARS, MERS

1. Introduction

Bats, with wide geographical distribution and capable of flying, while pangolins which harbor coronaviruses similar to the one that causes Covid-19, have been suspected to be the origins, which contain one of the largest groups of mammalian species and have been considered as natural hosts of a large number of diverse viruses such as lyssaviruses, paramyxoviruses and filoviruses (Smith and Wang, 2013). During ancient times, numerous novel coronaviruses have been discovered in a wide variety of bat similar to pangolins species throughout Asia, Europe, Africa and America (Drexler et al., 2014). Within the coronavirus genera alpha and beta-coronavirus, which mainly infect mammals, 7 out of the 15 currently assigned viral species have only been found in bats and pangolins (De Groot et al., 2012). It is suggested that bats and pangolins are major hosts for both alpha and beta-coronaviruses and perform significant function as the source of gene in the evolution of two coronavirus genera (Woo et al., 2012). Amidst the coronaviruses harbored by bats and pangolins, specific research interest has been drawn, as they exist to be associated with two high profile human disease outbreaks, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS).

Researchers in the current study focus on the growing cases of coronaviruses putatively linked to a zoonotic origin from bats and pangolins, represented by SARS coronavirus (SARS-CoV) and MERS coronavirus (MERS-CoV). Overview of current evidence for bat origin similar to pangolins of these two viruses and also discuss how the spillover events of coronavirus from animals to humans may have happened. Considering that bats have been known to harbor more coronaviruses than any other species, it is likely that SARS-CoV and MERS-CoV will not be the only bat and pangolin coronaviruses to jump among species and cause human infections.

Some studies have been conducted on the coronaviruses and some results stated that it originated from animals specifically bat, however; the origins remain a debate among scientists. Furthermore, few of these studies presented limited explanations about the origin of coronaviruses and none of them have detailed the history of the virus origins. Therefore, this study has collected, documented and compared data about coronaviruses.

* Correspondence: donmawolo26@gmail.com

This work is licensed under a Creative Commons Attribution 4.0 International License.
2. SARS and MERS first emergence

In 2002, SARS first emerged in Guangdong Province of southern China, as a novel clinical severe disease (termed “atypical pneumonia”) marked by fever, headache and subsequent onset of respiratory symptoms including cough, dyspnea, and pneumonia. Being highly transmissible among humans, SARS rapidly spread to Hong Kong and other provinces across China and then to other countries (Zhong et al., 2003; Chinese SARS Molecular Epidemiology Consortium, 2015). By July 2003, it had caused 8096 confirmed cases of infection in other countries, 774 (9.6%) of which were fatal. The second outbreak in 2004 only caused 4 infections without mortality nor further transmission (Song et al., 2005). The MERS epidemic emerged in the Kingdom of Saudi Arabia (KSA) since June 2012, with a similar clinical syndrome to SARS but seemingly less transmissible. In addition to respiratory illness, renal failure was identified in some severe cases (Bermingham et al., 2012; Zaki et al., 2012; World Health Organization, 2015). Unlike SARS which had numerous super-spreader events, most MERS cases were independent clusters and limited to countries in the Middle East, particularly in KSA. Limited MERS cases have been reported in African and European countries and the United States of America, but exclusively in individuals traveling back from the Middle East. Some patients were reported to have a history of contact with camels while many other cases lacked this epidemiological link (Bermingham et al., 2012; Zaki et al., 2012; World Health Organization, 2015). The MERS pandemic in the Republic of Korea in 2015 was caused by a single person who returned from travel in the Middle East. This made the Republic of Korea be home to the second-largest MERS epidemic with a total of 185 confirmed cases and 36 deaths (WHO, 2015; Korean Society of Infectious, 2015). By 18 August 2015, a total of 1413 laboratory-confirmed cases of MERS have been reported worldwide with a median age of 50 years, including 502 related deaths. The mortality of MERS (approximately 35%) is much higher than that of SARS (around 10%).

3. The representation of SARS-CoV and MERS-CoV species in the genus beta-coronavirus

3.1. Genomic structure and taxonomic classification

SARS-CoV and MERS-CoV share similar genome organization with other coronaviruses but display unique genomic structures and evolutionary lineages. The coronavirus genome possesses 6-to-7 major open reading frames (ORFs) in the characteristic gene order in the 5’ to 3’ direction: ORF1a and 1b which comprise two-thirds of the genome and encode the nonstructural polyproteins, and four ORFs downstream that encode structural proteins: spike protein (S), envelope protein (E), membrane protein (M) and nucleocapsid protein (N). Some coronaviruses have a hemagglutinin-esterase (HE) gene between ORF1b and S. Besides the coronavirus-conserved genes, the SARS-CoV genome contains a number of specific accessory genes including ORF3a, 3b, ORF6, ORF7a, 7b, ORF8a, 8b and 9b (Rota et al., 2003; Marra et al., 2003; Snijder et al., 2003). Comparably, MERS-CoV encodes five unique accessory genes, designated ORF3, ORF4a, ORF4b, ORF5, and ORF8b. None of these genes have been shown to be related to other known coronavirus genes at the time of discovery (Woo et al., 2012; Van Boheemen et al., 2012). MERS-CoV was found to have 75 and 77% amino acid (aa) sequence identity in 7 conserved replicate genes with two previously identified bat coronaviruses: BtCoV-HKU4 and BtCoV-HKU5. Based on the classification criteria of the International Committee on Taxonomy of Viruses (ICTV), SARS-CoV and MERS-CoV represent two novel distinct coronavirus species in the genus betacoronavirus (Table 1) (Zaki et al., 2012a: De Groot et al., 2013; Zaki, 2012). Members of beta-coronaviruses are separated into four lineages, A, B, C and D. SARS-CoV, and MERS-CoV are clustered in lineages B and C, respectively (De Groot et al., 2013).

3.2. Receptor usage

The S protein of coronaviruses is a surface-located trimeric glycoprotein consisting of two subunits: the N-terminal S1 subunit and the C-terminal S2 subunit. The S1 subunit specializes in recognizing and binding to the host cell receptor while the S2 region is responsible for membrane fusion. Compared with the S2, the S1 subunit shows much higher variability (Masters, 2006). Owing to its function of receptor binding, the variation in S protein defines in large part the tissue tropism and host range of different coronaviruses (Gallagher and Buchmeier, 2001).

---

Table 1. The analysis of bat and pangolin coronaviruses with other coronaviruses.

<table>
<thead>
<tr>
<th>Alphacoronavirus Section (A)</th>
<th>Betacoronavirus</th>
<th>Betacoronavirus lineage D</th>
<th>Rhinolophus sinicus BtCoV/Rs672 FJ588686</th>
<th>Rhinolophus sinicus BtCoV/HKU3 DQ022305</th>
<th>Betacoronavirus lineage D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miniopterus magnater BtCoV/1A NC_010437 (97)</td>
<td>Porcine PRCV/ISU-1 DQ811787 (100)</td>
<td>Roussettus aegyptiacus BtCoV/KY06 HQ728483</td>
<td>Rhinolophus macrotis BtCoV/Rm1 DQ412043</td>
<td>Rhinolophus blasii BtCoV/ BM48-31 NC_014470</td>
<td>Roussettus leschenaulti BtCoV/HKU9 EF_065513</td>
</tr>
<tr>
<td>Miniopterus pusillus BtCoV/1B NC_010436 (97)</td>
<td>Bovine BCoV/ENT NC_003045 (97)</td>
<td>Roussettus leschenaultia BtCoV/HKU9 EF065513</td>
<td>Rhinolophus sinicus BtCoV/ HKU3 DQ022305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miniopterus natalensis BtCoV/KY27 HQ_728484 (93)</td>
<td>Equine CoV/N99 NC_010327 (56)</td>
<td>Eidolon helvum BtCoV/KY24HQ728482</td>
<td>Rhinolophus sinicus BtCoV/ Rp3 HKU3 DQ071615</td>
<td>Gammaporonavirus</td>
<td></td>
</tr>
<tr>
<td>Miniopterus pusillus BtCoV/1B HKU8 NC_010438 (76)</td>
<td>Human CoV-OC43 NC_005147 (84)</td>
<td>Beta-coronavirus potential new line</td>
<td>Hipposideros commersoni BtCoV/ZBCoV HQ166910</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miniopterus spp. BtCoV/ HKU7 DQ_666339 (76)</td>
<td>Rat CoV/R1KF294370 (96)</td>
<td>Hipposideros spp. BtCoV/ GhanaKwan/20 FJ710047</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Porcine PEDV/MN KP468752 (100)</td>
<td>Mouse MHV-A59 NC_001846 (100)</td>
<td>Hipposideros spp. BtCoV/ GhanaKwan/20 FJ710047</td>
<td>Beta-coronavirus lineage A</td>
<td>Delta-coronavirus</td>
<td>Hipposideros Pomona BtCoV/HKU10 JQ989273</td>
</tr>
<tr>
<td>Porcine PEDV/AH2012 KC210145 (100)</td>
<td>Rat CoV/Parker NC_012936 (100)</td>
<td>Human CoV HKU1 NC_006577</td>
<td>Moorhen CMCoV/HKU11 FJ376620</td>
<td>Roussettus leschenaulti BtCoV/HKU10 NC_018871</td>
<td></td>
</tr>
<tr>
<td>Charophon spp. BtCoV/ KY22 HQ728486 (100)</td>
<td>Human CoV-HKU1 NC_006577 (100)</td>
<td>Betacoronavirus lineage B</td>
<td>Human CoV OC43 NC_005147</td>
<td>Bulbul BuCoV/HKU11 FJ376620</td>
<td>Human CoV NL63 NC_005831</td>
</tr>
<tr>
<td>Cardioderma cor BtCoV/ KY43 HQ728480 (100)</td>
<td>Human SARS-CoV/SZ3 AT304486</td>
<td></td>
<td>Bulbul BuCoV/HKU11 FJ376620</td>
<td></td>
<td>Human CoV 229E NC_002645</td>
</tr>
<tr>
<td>Hipposideros Pomona BtCoV/HKU10/Q989273 (100)</td>
<td>Human SARS-CoV/Tor2 NC_004718</td>
<td></td>
<td></td>
<td>Beta-coronavirus lineage C</td>
<td>Scotothipus kuhlii BtCoV/512/2005 NC_009657</td>
</tr>
<tr>
<td>Roussettus leschenaulti BtCoV/HKU10NC_018871 (100)</td>
<td></td>
<td></td>
<td>Rhinolophus ferrumequinum BtCoV/ YNLF KP886808</td>
<td>Camel MERS-CoV/KSA-CAMEL-376 KJ713299</td>
<td></td>
</tr>
<tr>
<td>Human CoV-NL63 NC_005831 (100)</td>
<td>Human MERS-CoV JX869059 (86)</td>
<td></td>
<td>Camel MERS-CoV/NRCE- HKU205KJ477102</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Section (B)
The phylogenetic table was constructed based on 816-nt partial RdRp sequences (Section A) and full-length spike protein sequences (Section B). Available sequences were retrieved from GenBank and aligned using ClustalW. The alignment was used to construct a tree by MEGA (Version 5.1) with the neighbor-joining statistical method. Bootstrap values were calculated from 1000 replicates (values ≥50 are shown). Bat coronaviruses are written in bold and named following bat species, plus BtCoV, strain name, and GenBank accession number.

**Table 1.** (Continued).

<table>
<thead>
<tr>
<th>Bat Species</th>
<th>MERS-CoV or SARSCoV Strain</th>
<th>Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Carollia perspicillata</em> BtCoV/1FY2BA</td>
<td>Camels MERS-CoV-NRCE-HKU205KJ477102 (100)</td>
<td>EU769557</td>
</tr>
<tr>
<td></td>
<td>Rhinolophus sinicus BtCoV/Rs3367 KC881006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tylonycteris pachypus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human SARS-CoV/Tor2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NC_004718</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Myotis lucifugus</em> BtCoV/CDPHE15/USA NC_022103 (100)</td>
<td>Camels MERS-CoV-KSA-CAMEL-376 KJ713299 (990)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus sinicus BtCoV/RsWIV1KF367457</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pipiterillus abramus BtCoV/HKU5 NC_009020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Civet SARS-CoV/SZ3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AY304486</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Myotis daubentonii</em> BtCoV/ NM98-62 GU190216 (57)</td>
<td>Neoromica capensis BtCoV/NeoCoV KC869678 (99)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus sinicus BtCoV/Rp3DQ071615</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilio superans BtCoV/SC2013 KJ473821</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus sinicus BtCoV/Rs3367 KC881006</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Scotophilus kuhlii</em> BtCoV/512/2005 NC_009657 (57)</td>
<td>Pipistrellus pygmaeus BtCoV/8-724 KC243390 (56)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus sinicus BtCoV/Rs672 FJ588686</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Erinaceus europaeus EriCoV NC_022643</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus sinicus BtCoV/W1V1 KF367457</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hipposideros spp.</em> BtCoV/ Ghanakwam/19 FJ710046 (58)</td>
<td>Tylonycteris pachypus BtCoV/HKU4 NC_009019 (81)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chaerephon plicata BtCoV/Cp JX993988</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neoromica capensis BtCoV/Neo KC869678</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus sinicus BtCoV/W1V1 KF367457</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Human CoV-229E</em> NC_002645</td>
<td>Erinaceus europaeus EriCoV NC_022643 (66)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus Ferrumequinum BtCoV/YNFL 31C KP886808</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus macrotis BtCoV/Rm1 DQ412043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus sinicus BtCoV/RsHOC014 KC881005</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chaerephon pumilus</em> BtCoV/KenyaK41 HQ728481</td>
<td>Rhinolophus Ferrumequinum BtCoV/Rf1 DQ412042</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus ferrumequinum BtCoV/Rf1DQ412042</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus pusillus BtCoV/Rp JX993987</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Feline FIPV/79-1146 AY994055</em></td>
<td>Chaerephon plicata BtCoV/Cp JX993988</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophus pusillus BtCoV/Rp JX993987</td>
<td></td>
</tr>
</tbody>
</table>
Angiotensin-converting enzyme 2 (ACE2) was identified to be the functional receptor of SARS-CoV (Li et al., 2003; Hamming et al., 2004; Ding et al., 2004). A 193 aa fragment (aa 318–510) of SARS-CoV S protein was demonstrated to bind ACE2 more efficiently than the full S1 domain and was defined as the receptor-binding domain (RBD) of SARS-CoV (Wong et al., 2004). A loop subdomain (aa 424–494) that directly contacts with ACE2 was further identified as the receptor-binding motif (RBM) by crystal structure analysis (Li et al., 2005). In the RBM, several aa residues were found to be critical for receptor binding and changes in these key residues resulted in different binding efficiency among different SARS-CoV isolates (Li et al., 2005; Qu et al., 2005; Li et al., 2005). Dipeptidyl peptidase 4 (DPP4, also known as CD26) was identified as a functional receptor for MERS-CoV (Raj et al., 2003) and is relatively conserved among mammalian species. Published results indicated that MERS-CoV can infect and replicate in most cell lines derived from a human, nonhuman primate, bat, swine, goat, horse, rabbit, civet, and camel, but not from mice, hamster, dog, ferret, and cat (Raj et al., 2003; Barlan et al., 2014; De Wit et al., 2013; Coleman et al., 2014; Eckerle et al., 2014; Chanet et al., 2013; van Doremalen et al., 2014; Haagmans et al., 2015). DPP4 from camel, goat, cow, and sheep can be also recognized by MERS-CoV and can support MERS-CoV replication (Barlan et al., 2014; Van Doremalen et al., 2014). Resolved crystal structures demonstrate that DPP4-recognizing RBD is localized to the S1 C-terminal portion of S protein of MERS-CoV (Lu et al., 2013; Wang et al., 2013; Chenet al., 2013). The RBD of MERS-CoV consists of ~240 residues, spanning aa 367–606, which fold into a structure consisting of two subdomains, the core subdomain, and the external subdomain. The core subdomain of MERS-CoV RBD is structurally similar to that of the SARS-CoV RBD, but the external subdomain (also named as RBM) is different from that of the SARS-CoV (Lu et al., 2013; Wang et al., 2013; Chenet al., 2013).

4. The origins of bat and pangolins of SARS-CoV

Civets are intermediate and transmission host of SARS-CoV. Epidemiological survey showed that early cases of SARS-CoV in 2002–2003 and all 4 cases in 2003–2004 had a history of animal contact through animal trade in wet markets and restaurants where live animals were kept in Guangdong, China. Molecular detection and virus isolation studies suggested that the pandemic-causing SARS-CoV originated from traded civets in wet markets. This was indirectly confirmed by the massive culling of market civets, which was believed to play a major role in efficiently containing the SARS pandemics and no further SARS case was reported after 2004 (Guan et al., 2003; Centers for Disease and Prevention Control, 2003; Xu et al., 2004). However, subsequent extensive epidemiology studies did not find SARS-CoV in farmed or wild-caught civets, indicating that another animal(s) was involved in SARS-CoV transmission in the animal market or other trading activities and civets are unlikely the natural reservoir of SARS-CoV (Kan et al., 2005; Poon et al., 2005; Wu et al., 2005).

4.1. Observation of diverse SARS-like coronaviruses in bats and pangolins

Several years before the outbreak of SARS, two other zoonotic viruses, Nipah virus and Hendra virus emerged in Asia and Australia, and they were both known to be originated from bats (Halpin et al., 2000; Yob et al., 2001). These findings have led scientists to consider bats and pangolins in the search of reservoirs of SARS-CoV. In 2005, a breakthrough was made as two independent research groups reported, almost simultaneously, the discovery of novel coronaviruses related to SARS-CoV in horseshoe bats and pangolins (in the genus Rhinolophus and Manis) in China, which were termed SARS-like coronavirus (SL-CoV) (Li et al., 2005; Lau et al., 2005). These bat SL-CoVs from both mainland China and Hong Kong manifested a genome sequence identity of 88%–90% among themselves and 87%–92% identity to human or civet SARS-CoV isolates. The unique set of ORFs exclusively found in SARS-CoV was also present in bat and pangolin SL-CoVs, demonstrating the close phylogenetic relationship between SARS-CoV and SL-CoV. The discovery of bat SL-CoV boosted researchers’ interest in coronavirus surveillance studies in bats. In the following years, SL-CoV RNA was detected in Rhinolophus species of a wider geographic range in China. The provinces or regions where SL-CoV-positive bats were captured included Hong Kong, Guangxi, Hubei, Shandong, Guizhou, Shaanxi, and Yunnan (Tang et al., 2006; Woo et al., 2006; Yuan et al., 2010; Ge et al., 2013). 7 conserved replicate domains in orf1ab of these SL-CoVs found in China were compared with those of SARS-CoV (Table 2). They all shared higher than 95% aa sequence identity with SARS-CoV in the concatenated domains and therefore can be considered to belong to SARS-CoV species (King et al., 2012). SL-CoVs were also discovered in rhinolophids from Countries in Europe (Drexler et al., 2010; Riihertic et al., 2010; Balboni et al., 2011). These European SL-CoVs exhibited significant genetic variation from Chinese isolates. The strain BM48-31 from Rhinolophus blasii in Bulgaria was highly divergent from Chinese isolates, displaying major sequence differences in several genes including ORF3b and ORF6 and lacking the coding region of ORF8 in its genome (Drexler et al., 2010). In Africa, novel beta-coronaviruses related to SARS-CoV have been detected in Hipposideros and Chaerophon species from Africa. However, compared with Asian and European SL-CoVs, these viruses of nonrhinolophid
Table 2. Comparison of bat coronaviruses with SARS-CoV or MERS-CoV in conserved replicase domains and structural proteins.

<table>
<thead>
<tr>
<th>CoV Strain</th>
<th>Bat and pangolins species</th>
<th>Regions</th>
<th>Percent of amino acid sequence identity with SARS-CoV or MERS-CoV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ADRP</td>
</tr>
<tr>
<td>HKU3</td>
<td>Rhinolophus sinicus</td>
<td>China</td>
<td>92.0</td>
</tr>
<tr>
<td>Rp3</td>
<td>Rhinolophus sinicus</td>
<td>China</td>
<td>95.4</td>
</tr>
<tr>
<td>Rm1</td>
<td>Rhinolophus macrotis</td>
<td>China</td>
<td>91.0</td>
</tr>
<tr>
<td>RI1</td>
<td>Rhinolophus ferrumequinum</td>
<td>China</td>
<td>92.3</td>
</tr>
<tr>
<td>Rs672</td>
<td>Rhinolophus sinicus</td>
<td>China</td>
<td>97.0</td>
</tr>
<tr>
<td>Rs3367</td>
<td>Rhinolophus sinicus</td>
<td>China</td>
<td>97.0</td>
</tr>
<tr>
<td>RsSHC014</td>
<td>Rhinolophus sinicus</td>
<td>China</td>
<td>96.9</td>
</tr>
<tr>
<td>WIV1</td>
<td>Rhinolophus sinicus</td>
<td>China</td>
<td>97.0</td>
</tr>
<tr>
<td>Cp/Yunnan</td>
<td>Chaerephon plicata</td>
<td>China</td>
<td>97.6</td>
</tr>
<tr>
<td>Rp/Gansu</td>
<td>Rhinolophus pusillus</td>
<td>China</td>
<td>93.5</td>
</tr>
<tr>
<td>YNLF_31C</td>
<td>Rhinolophus ferrumquinum</td>
<td>China</td>
<td>97.2</td>
</tr>
<tr>
<td>BM48-31</td>
<td>Rhinolophus blasii</td>
<td>Bulgaria</td>
<td>76.8</td>
</tr>
<tr>
<td>HKU4-1</td>
<td>Tylonycteris pachypus</td>
<td>China</td>
<td>81</td>
</tr>
<tr>
<td>HKU5-1</td>
<td>Pipistrellus abramus</td>
<td>China</td>
<td>55.5</td>
</tr>
<tr>
<td>NeoCoV</td>
<td>Neoromia capensis</td>
<td>South Africa</td>
<td>86.7</td>
</tr>
<tr>
<td>SC2013</td>
<td>Vespertilio superans</td>
<td>China</td>
<td>53.5</td>
</tr>
</tbody>
</table>

Calculated with MEGA5.1 using a pairwise deletion option; Bat SL-covs are listed in the upper part of the table while camel MERS-CoV and pangolins covs related to MERS-cov in the lower part seven domains were series connected and calculated with MEGA5.1 using a pairwise deletion option ADRP, ADP-ribose 1-phosphatase; 3clpro, coronavirus NSP5 protease; rdrp RNA-dependent RNA polymerase; Hel, helicase; exon, exoribonuclease; nendou, endoribonuclease; OMT, 2’-O-methyltransferase genbank accession numbers: Tor2, NC_004718; HKU3, DQ022305; Rp3, DQ071615; Rm1, DQ412043; RI1, DQ412042; Rs672, FJ588686; Rs3367, KC881006; rsshc014, KC881005; WIV1, KF367457; Cp/Yunnan2011, JX993988; Rp/Shaanxi2011, JX993987; YNLF_31C, KP886808; EMC/2012, JX869059; HKU5-1, NC_009020; HKU4-1, NC_009019; betacov/SC2013, KJ473821; neocov, KC869678.
4.2. The ancestor of SARS-CoV in bats and pangolins
Although it is mentioned that bat and pangolins SL-CoVs showed high sequence identity to SARS-CoV, two deletions were present in the RBM of their S proteins (Li et al., 2005; Lau et al., 2005). The differences in RBM substantially changed receptor usage. In a study using an HIV-based pseudovirus system and cell lines expressing human, civet, and horseshoe bat ACE2 molecules, the bat and pangolin SL-CoV Rp3 S protein demonstrated its inability to use ACE2 as cell receptor (Ren et al., 2008). However, the chimeric Rp3 S protein carrying the RBD of SARS-CoV S protein was conferred the capability of cell entry via human ACE2 (Ren et al., 2008). These results suggested that bat and pangolin SL-CoVs such as Rp3 were unlikely to cause human infection. Therefore, they may not be considered as the direct progenitor of SARS-CoV. Besides, the theory of bat and pangolin origin of SARS-CoV lacked powerful support due to the failure of direct isolation of SL-CoV from bats, despite numerous trials by our group as well as many others around the world. During our longitudinal surveillance at a Rhinolophus sinicus colony in Yunnan Province over the years, a major breakthrough came in 2013 when diverse SLCoVs were discovered in the single colony (Ge et al., 2013). In this colony, there were at least 7 different strains related to SARS-CoV, HKU3, Rs672 or Rf1, based on analysis of the region corresponding to SARS-CoV RBD. Intriguingly, unlike all previously described SL-CoVs, two strains, designated Rs3367 and RsSHC014, did not contain the deletions in this region. Rs3367 showed a particularly high sequence identity to SARS-CoV in RB and was identical to SARS-CoV in several key amino acid residues known to be important for receptor binding (Ge et al., 2013). Whole genome sequencing revealed that Rs3367 and RsSHC014 shared more than 95% genome sequence identity with human and civet SARS-CoV, which was remarkably higher than that of any other bat SL-CoV (76 to 92%). Regarding individual genes, the amino acid sequence identity between Rs3367 or RsSHC014 and SARS-CoV was higher than 96% in ORF1a, 1b, 3a, 3b, E, M and N genes (Ge et al., 2013). Most importantly, a live SL-CoV was isolated for the first time from bat fecal samples (Ge et al., 2013). This virus, termed WIV1, had almost identical sequences (99.9%) to Rs3367 and was demonstrated to use ACE2 molecules from humans, civets and Chinese horseshoe bats and pangolins for cell entry. It also displayed infectivity in cell lines from a broad range of species including human, pig, bat, and pangolin. Furthermore, the close relatedness between WIV1 and SARS-CoV was confirmed by the neutralization effect of convalescent SARS patient sera on WIV1 (Ge et al., 2013). The isolation of a bat SLCoV genetically closely resembling SARS-CoV and having a functional S protein capable of using the same ACE2 receptor as SARS-CoV provided robust and conclusive evidence for the bat origin of SARS-CoV.

4.3. Possible origin of SARS-CoV from recombination of different SL-CoVs
Despite the fact that Rs3367 or WIV1 is unprecedentedly close to SARS-CoV in terms of RB region and genome identity, still, there are gaps between them and the immediate ancestor of SARS-CoV. ORF8 is a highly variable gene and remarkable differences can be observed among SARS-CoVs and SL-CoVs of different host origins. Isolates from civets and from early phase of the 2002/2003 pandemic contained a single long ORF8, while in the human SARS-CoV isolates from the middle and late phase of the pandemic the ORF8 was disrupted into two ORFs, ORF8a and ORF8b, as a result of the acquisition of a 29-nt deletion after interspecies transmission to humans (Songet al., 2005; Guanet al., 2003; Quan et al., 2010). The SL-CoVs from Rhinolophus sinicus, including Rs3367, however, had a single ORF8 with only 32%–33% amino acid identities to that of civet SARS-CoV. In contrast, the ORF8 of two novel SL-CoV strains recently reported in Yunnan from another rhinolophid species, Rhinolophus ferrumequinum, exhibited exceptionally high (81.3%) amino acid identity to civet SARS-CoV SZ3 (Lau et al., 2015). This is consistent with isolate Rf1, an SL-CoV reported earlier from R. ferrumequinum in Hubei Province, China of which the ORF8 shared 80.4% amino acid identity to SZ3 (Li et al., 2005). Potential recombination sites were identified around the ORF8 region between SL-CoVs from R. sinicus and R. ferrumequinum and it has been suggested that the ancestor of civet SARS-CoV probably acquired ORF8 from R. ferrumequinum SL-CoVs by recombination (Lau et al., 2015).

4.4. Animal origins of MERS-CoV
As with SARS-CoV, most early MERS cases had contact history with animals, e.g., dromedary camels (Albarrak et al., 2012; Health Protection Agency, 2013). MERS-CoV RNA was detected in camels from Saudi Arabia, Qatar, and Egypt and showed high similarities (>99%) to human MERS-CoV in genomic sequences (Memish et al., 2014; Chu et al., 2014; Haagmans et al., 2014; Briese et al., 2014; Yusof et al., 2015; Annan et al., 2015). Serological evidence further confirmed a high prevalence of MERS-CoV infections in camels in the Middle East (Hemida et al., 2013; Reusken et al., 2013; Meyer et al., 2014; Nowotny N et al., 2014; Alagaili et al., 2014; Reusken et al., 2013), Africa (Corman et al., 2014; Reusken et al., 2014; Muller et al., 2014) and Europe (Reusken et al., 2014).
neutralization antibodies in camels could be traced back to 1983 (Reusken et al., 2014; Muller et al., 2014). These results strongly suggested that MERS-CoV infection in humans was transmitted through close contact with infected camels (Memish et al., 2014; Alagaili et al., 2014; Azhar et al., 2014a; Azhar et al., 2014b).

4.5. Bat and pangolins viruses related to MERS-CoV

Prior to the emergence of MERS-CoV, a group of bat and pangolin coronaviruses had been reported including Tylonycteris bat and pangolin coronavirus HKU4 (Bt/PgCoV-HKU4) in Tylonycteris bats and Pipistrellus bat coronavirus HKU5 (BtCoVHKU5) in Pipistrellus bats and pangolin in China (Tang et al., 2006; Woo et al., 2007; Lau et al., 2013), E.isa/ M/Spain/2007 in Eptesicus isabellinus bats in Spain (Falcon et al., 2011) and N.noc/ VM366/2008/NLD in Pipistrellus pipistrellus bats in the Netherlands (Reusken et al., 2010). Based on genomic sequence analysis, these bat coronaviruses were grouped into lineage C of the genus beta-coronavirus. After the outbreak of MERS, MERS-CoV related coronaviruses were found in more bat and pangolin species and countries (De Benedictis et al., 2014; Anthonyet al., 2014; Annan et al., 2013; Wacharapluesadee et al., 2013; Corman et al., 2014; Ithete et al., 2014; Memish et al., 2013; Yang et al., 2014; Lelli et al., 2013). Among these viruses, full-length or near full-length genomes of Bt/PgCoV-HKU4, Bt/PgCoV-HKU5, SC2013 and NeoCoV have been characterized. By genomic analysis of lineage C beta-coronaviruses, MERS-CoV derived from camels show high similarities to human MERS-CoV with >99.5% nt identities, confirming that the human and camel isolates belong to the same coronavirus species. Bat and pangolin HKU4, HKU5, NeoCoV, and SC2013, shared 69.8%, 70%, 85.6% and 75.6% nt identities with MERS-CoV at the genomic level, respectively. Seven conserved replicase domains in orf1ab of MERS-CoV related viruses were compared with MERS-CoV (Table 2). The concatenated translated domains of NeoCoV shared 95% aa sequence identity with MERS-CoV and it could be classified as the same MERS-CoV species (King et al., 2013). Other bat and pangolin coronaviruses, HKU4, HKU5, and SC2013, could be considered as different coronavirus species. The most recent ancestor analysis speculated that MERS-CoV may have jumped from bats to camels approximately 20 years ago in Africa, with camels then being imported into the Arabian Peninsula (Corman et al., 2014), while HKU5 and MERS-CoV may have diverged from their common ancestor about 400 to 500 years ago (Lau et al., 2013). Although NeoCoV is closer to MERS-CoV than other bat coronaviruses at the genomic level, the phylogenetic analysis of the spike protein showed that HKU4 is the most closely related to MERS-CoV among all currently known bat and pangolin coronaviruses, sharing 67% sequence identity (Tables 1 and 2). This is correlated with the capability of HKU4 of using DPP4 as its functional receptor. However, HKU4 preferred bat DPP4 over human DPP4, whereas MERS-CoV showed the opposite trend (Yang et al., 2014). It was suggested that MERS-CoV ancestors had been circulating in bats and pangolins for a very long time. MERS-CoV has evolved to adapt to the use of the human receptor and the DPP4-recognizing bat coronaviruses like HKU4 may follow up, thereby posing a serious risk to human health (Yang et al., 2014; Cui et al., 2013).

4.6. Transmission of SARS-CoV

During December, the Chinese health authority reported an outbreak of serious pneumonia disease in Wuhan, China. The causative agent was soon identified as a novel coronavirus (Wu et al., 2020), which was later named SARS-CoV-2. Case numbers grew rapidly from December 27 – the Chinese annual festival, 3,090,445 globally as of 30 April 2020 (Lu et al., 2020), leading to the declaration of a public health emergency, and later a pandemic, by the WHO (World Health Organization). Many of the early cases were linked to the Huanan seafood market in Wuhan city, Hubei province, from where the probable zoonotic source is speculated to originate (WHO 2019). Currently, only environmental samples taken from the market have been reported to be positive for SARS-CoV-2 by the Chinese Center for Disease Control and Prevention (Cohen et al., 2020). However, as similar wet markets were implicated in the SARS outbreak of 2002–2003 (Wang et al., 2005), it seems likely that wild animals were also involved in the emergence of SARS-CoV-2. Indeed, many mammalian species were available for purchase in the Huanan seafood market before the outbreak (Cohen et al., 2020). Unfortunately, because the market was cleared soon after the outbreak began, determining the source virus in the animal population from the market is challenging. During December, the Chinese health authority reported an outbreak of serious pneumonia disease in Wuhan, Hubei province, from where the probable zoonotic source is speculated to originate (WHO 2019). Currently, only environmental samples taken from the market have been reported to be positive for SARS-CoV-2 by the Chinese Center for Disease Control and Prevention (Cohen et al., 2020). However, as similar wet markets were implicated in the SARS outbreak of 2002–2003 (Wang et al., 2005), it seems likely that wild animals were also involved in the emergence of SARS-CoV-2. Indeed, many mammalian species were available for purchase in the Huanan seafood market before the outbreak (Cohen et al., 2020). Unfortunately, because the market was cleared soon after the outbreak began, determining the source virus in the animal population from the market is challenging. Although a coronavirus that is closely related to SARS-CoV-2, which was sampled from a Rhinolophus affinis bat in Yunnan in 2013, has now been identified (Zhou et al., 2020), similar viruses have not yet been detected in other wildlife species. The current researcher revealed SARS-CoV-2-related viruses in pangolins smuggled into southern China.

The observation of putative recombination signals between the pangolin coronaviruses, bat coronavirus RaTG13, and human SARS-CoV-2. Particularly, the SARS-CoV-2 exhibits very high sequence similarities reported in Guangdong pangolin coronaviruses in the receptor-binding domain (RBD) (97.4% amino acid) (Lam et al., 2020) even though it is most closely related to bat coronavirus RaTG13 in the remainder of the viral genome. Indeed, the Guangdong pangolin coronaviruses and SARS-CoV-2 possess identical amino acids at the five critical residues of the RBD, whereas RaTG13 only...
shares one amino acid with SARS-CoV-2 (residue 442, according to the numbering of the human SARS-CoV) (Wan et al., 2020; Lam et al., 2020) and these latter two viruses have only 89.2% amino acid similarity in the RBD. Notably, a phylogenetic analysis of synonymous sites only from the RBD revealed that the topological position of the Guangdong pangolin is consistent with that of the remainder of the viral genome, rather than being the closest relative of SARS-CoV-2 (Lam et al., 2020). Therefore, the amino acid similarity between the RBD of the Guangdong pangolin coronaviruses and SARS-CoV-2 may be due to selectively mediated convergent evolution rather than recombination. This observation is consistent with the fact that the sequence similarity of ACE2 is higher between humans and pangolins than between humans and bats (Lam et al., 2020). The occurrence of recombination and/or convergent evolution further highlights the role that intermediate animal hosts have in the emergence of viruses that can infect humans. However, all of the pangolin coronaviruses identified to date lack the insertion of a polybasic (furin-like) S1/S2 cleavage site in the spike protein that distinguishes human SARS-CoV-2 from related beta coronaviruses (including RaTG13) (Coutard et al., 2020) and that may have helped to facilitate the emergence and rapid spread of SARS-CoV-2 through human populations.

To our knowledge, pangolins are the only mammals in addition to bats that have been documented to be infected by a SARS-CoV-2-related coronavirus. Notably, two lineages of coronaviruses are found in pangolins that were independently sampled in different areas of China and that both are also related to SARS-CoV-2. This suggests that these animals may be important hosts for these viruses, which is surprising as pangolins are solitary animals that have relatively small population sizes, reflecting their endangered status (Heinrich et al., 2017). Indeed, based on the current data it cannot be excluded that pangolins acquired their SARS-CoV-2-related viruses similarly from bats or another animal host. Therefore, their role in the emergence of human SARS-CoV-2 needs further attention. In this context, it is noteworthy that both lineages of pangolin coronaviruses were obtained from pangolins, which originated from Southeast Asia, and that there is a marked lack of knowledge of the viral diversity maintained by this species in regions in which it is indigenous. Furthermore, the extent of virus transmission in pangolin populations should be investigated further. However, the repeated occurrence of infections with SARS-CoV-2-related coronaviruses in Guangxi and Guangdong provinces suggests that this animal may have an important role in the community ecology of coronaviruses.

Coronaviruses, including those related to SARS-CoV-2, are present in many wild mammals in Asia (Wang et al., 2017). Although the epidemiology, pathogenicity, interspecies infectivity and transmissibility of coronaviruses in pangolins remains to be studied, the data presented here strongly suggest that handling these animals requires considerable caution and their sale in wet markets should be strictly prohibited. Further surveillance of pangolins in their natural environment in China and Southeast Asia are necessary to understand their role in the emergence of coronaviruses and the risk of future zoonotic transmissions.

4.7. Bat coronaviruses and human coronavirus 229E

The (HCoV-229E) and NL63 (HCoV-NL63) was found in the 1960s and causes comparatively mild common colds worldwide (Reed, 1984). A bat coronavirus detected in Hipposideros caffer ruber in Ghana termed Hipposideros/GhanaKwam/19/2008 was genetically related to HCoV-229E. Its RdRp fragment shared 92 % nucleotide sequence identity with HCoV-229E and they were predicted to share a most recent common ancestor (MRCA) only 200 years ago (Pfaffler et al., 2009). A recent study characterized more 229E-related coronaviruses discovered in hipposiderid bats from Ghana on full genome level. These bat coronaviruses were more diversified and formed a single viral species with HCoV-229E. Interestingly, phylogenetic analysis revealed the intermediate position of a 229E-related alpaca virus between bat and human viruses. These findings suggested the ancestral origin of HCoV-229E in hipposiderid bats and the role of camelids as potential intermediate hosts was hypothesized (Corman et al., 2015). HCoV-NL63 was first isolated from babies suffering from pneumonia and bronchiolitis in 2004 (Fouchier et al., 2004). To date, HCoV-NL63 has been found worldwide with up to 9.3% detection rate in hospitalized respiratory tract samples (Fielding et al., 2011). In 2010, a bat coronavirus termed ARCoV2 (Appalachian Ridge CoV) detected in North American tricolored bat (Perimyotis subflavus) in the US showed a close relationship with HCoV-NL63. The MRCA for HCoV-NL63 and ARCoV2 was predicted to have existed 563 to 822 years ago (Donaldson EFet al., 2010; Huynh Jet al; 2012). Further analysis indicated that HCoV-NL63 can replicate in cell lines derived from the lungs of tricolored bats (Huynh Jet al; 2012). These results suggest that prototypes of HCoVNL63 may also exist in bats and there may also be a bat origin of this human coronavirus.

5. Conclusion

The study documented that if not all, currently circulating alpha coronaviruses and beta-coronaviruses in different mammals are evolutionally linked to ancestral coronaviruses originated from bats and pangolins. Different species of rhinolophid bats and Manis pangolins in China carry genetically diverse SARS-like coronaviruses, some of which are direct ancestors of SARS-CoV and hence have
the potential to cause the direct interspecies transmission to humans.

Meanwhile, different coronavirus species closely related to MERS-CoV are circulating in bats and pangolins. Bats and pangolins might likely be the natural reservoirs of MERS-CoV or an ancestral MERS-like CoV. It is hypothesized that bat MERS-like CoV jumped to camels or some other as yet unidentified animal sources some years ago. The virus evolved and adapted with accumulating mutations in camels and then was transmitted to humans very recently. It took almost a decade from the first discovery of SL-CoV in bats and pangolins to the final isolation of the SARS-CoV ancestral virus from bats, so continuing surveillance is vital to uncover the origin of MERS-CoV and bats should certainly be a priority of research. Besides, as the spike protein and host receptor are key factors of cross-species transmission of coronaviruses, characterization of the receptor and key binding sites of the spike protein will be important in estimating host tropism of bat coronaviruses and predicting spillover risk. With human activity increasingly overlapping the habitats of bats and pangolins, disease outbreaks resulting from the spillover of bat coronaviruses will continue to occur in the future despite the fact that direct transmission of bat and pangolin coronaviruses to humans appears to be rare. To prevent the next outbreak, it is necessary to maintain vigilance in long-term coronavirus surveillance studies in bats and pangolins as well as in other wildlife and livestock.

**Acknowledgments**

The authors extend heartfelt thanks and appreciation to the National Public Health Institute of Liberia for the numerous assistance. Special thanks to Dr. Mosoka P. Fallah and Mr. Bode Ireti Shobayo for his technical advice during the research.

**Conflict of interest**

The authors declare no conflict of interest.

**Funding**

The researchers receive no funding for this work.

**Data accessibility**

All data are provided in the full section of this manuscript.

**Contribution of authors**

The mentioned authors performed conceptualization, formal analysis, collected the samples, review, wrote the original-draft, review, and editing. James Blackar Mawolo is the corresponding author. All authors contributed to this manuscript revision, read and approved the submitted version.

**References**


