Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Evolution of SARS Coronavirus and the Relevance of Modern Molecular Epidemiology

26

Z. Shi [1], *L.-F. Wang* [2]
[1]Chinese Academy of Sciences (CAS), Wuhan, China; [2]Duke-NUS Medical School, Singapore, Singapore

1. A Brief History of SARS

As outlined in Table 26.1, the first reported case of "atypical pneumonia," now known as Severe Acute Respiratory Syndrome or SARS, occurred in Guangzhou, Guangdong province, China, on November 16, 2002. Before the end of February 2003, a total of 11 index cases occurred independently in nine cities of Guangdong Province, which forms the early phase of the SARS epidemic.[1] These index cases spread the virus to their close relatives and hospital staffs and provided the early demonstration of the respiratory transmission mode of the disease. The clinical symptoms of SARS are nonspecific. The index cases all began to have fever higher than 38°C and displayed common respiratory symptoms, such as cough, headache, and shortness of breath.

The dynamics of the outbreak was largely shaped by the presence of the so-called super spread event (SSE), in which a single patient was shown to spread the virus to a large number of contacts.[1] It is the SSEs that triggered the large scale of SARS pandemic in China. The first SSE patient is a businessman specialized in fishery wholesale. He was treated in three hospitals from January 30, 2003 to February 10, 2003 and along the way infected at least 78 other individuals including hospital staffs, patients, and close relatives and friends.[1] The second SSE individual, who caused the major spread of the disease out of Guangdong, was a business lady, native of Shanxi province. She went to Guangdong for business in late February and become sick while traveling. She went back to her home province and infected eight family members as well as five hospital staffs. The spread continued to Beijing when she decided to seek better treatment in Beijing.[1,2]

The beginning of the global transmission occurred in Metropole Hotel of Hong Kong where a professor of nephrology from a Guangdong hospital stayed during a private visit. Without knowing, the urologist was infected with SARS-CoV a few days before he traveled to Hong Kong. It is later found that he spread the virus to at least 15 other persons in the hotel and in the hospital where he was treated. Among them, five of the hotel contacts continued their international journeys and further transmitted the disease to Vietnam, Singapore, Canada, and other countries. This marks the true beginning of a disastrous worldwide pandemic (http://www.who.int/csr/sars/en/).

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00026-3
Copyright © 2017 Elsevier Inc. All rights reserved.

Table 26.1 Chronological Events of the SARS Outbreaks

Date	Event
November 16, 2002	The first recognized SARS patient, in Foshan, Guangdong province, China
November 16, 2002 to March 10, 2003	11 independent index cases in Foshan, Heyuan, Jiangmen, Zhongshan, Shunde, Guanzhou, Zhaoqing, Shenzhen, Dongguan, China, resulting in more than 50 secondary infections
January 22, 2003	SARS spreading in Guangdong province
March 22, 2003	SARS spreading to Shanxi and Beijing
February 21, 2003	SARS spreading to Hong Kong, marking the beginning of the global pandemic
February 28, 2003	SARS spreading to Vietnam
March 12, 2003	WHO global travel alert of the SARS pandemic
March 14, 2003	SARS spreading to Canada
March 6, 2003	SARS spreading to Singapore
March 17, 2003	WHO established a 9-nation/11-institute international laboratory network
March 24, 2003	Coronavirus was isolated from SARS patient
April 4, 2003	SARS spreading to Philippines
April 12, 2003	Full-length genome of SARS-CoV determined
April 17, 2003	The international laboratory network announced conclusive identification of SARS-CoV as the causative agent
May 23, 2003	Detected SARS coronavirus in market animals
July 5, 2003	WHO removed the last region from the affected list, effectively marking the end of the outbreak
August 7, 2003	WHO reported a total of 8096 cases and 774 death covering the major 2002−2003 outbreaks
September 2003 to April 2004	Outbreaks caused by laboratory incidents in Singapore, Taiwan, and Beijing
December 16, 2003 to January 8, 2004	Four independent SARS cases in Guangdong, causing mild disease with no death

WHO played a key role in the investigation and control of the SARS outbreak from the very beginning. For the first time in history, WHO issued a global travel alert on March 12, 2003, which greatly reduced the rate of long-distance transmission of the disease. On March 17, 2003 WHO established a 9-nation/11-institute SARS network that played a major role in the rapid identification of the causative agent and development of diagnostic tests. Thanks to the international effort co-coordinated by WHO, the SARS

outbreaks were effectively under control by July 5, 2003. This was the first powerful demonstration of the kind of devastation a new infectious disease can cause worldwide and the effectiveness of an international organization when it is running at its peak.

Following the major SARS outbreaks of 2003–2004, there were several minor outbreaks with much smaller impacts. Between December 2003 and January 2004, four independent SARS cases were reported in Guangdong, and none of them led to fetal infection or widespread transmission. Subsequent epidemiological tracing revealed that all cases could be linked to civet trading activities.[3] In addition, there were three laboratory outbreaks in September 2003, December 2003, and April 2004 in Singapore, Taiwan, and Beijing, respectively. The most severe outbreak was associated with the incident in Beijing that resulted in a total of nine infection cases with one death. None of the other two laboratory infections resulted in further spread of the virus.[4]

2. SARS Coronavirus

Rapid identification of causative agent in an outbreak caused by unknown pathogen is the key for an effective response. However, in the case of SARS outbreak, this was not the case. Due to the association of nonspecific clinical symptoms associated with SARS patients, several pathogens were initially "identified" as the potential causes of SARS, which included *Chlamydia*, influenza virus, and paramyxovirus.[5] The confusion continued until March, 2003 when three laboratories independently confirmed that a previously unknown coronavirus was the most likely etiological agent of SARS.[6–8]

Coronaviruses are enveloped viruses with the largest single-stranded, positive-sense RNA genomes currently known, ranging in size from 27 to nearly 32 kb in length. Coronaviruses can infect and cause disease in a broad array of avian and mammal species, including humans. The name "coronavirus" is derived from the Greek word, meaning crown, as the virus envelope appears under electron microscopy to be crowned by a characteristic ring of small bulbous structures. Within the virion, the ssRNA genome is encased in a helical nucleocapsid composed of many copies of the nucleocapsid (N) protein. The lipid bilayer envelope contains three proteins: envelope (E) and membrane (M) protein, which coordinate virion assembly and release, and the large spike (S) protein, which confers the virus's characteristic corona shape as well as serves as the principal mediator of host cell attachment and entry via virus- and host-specific cell receptors. The size of the SARS-CoV viral particle is approximately 80–90 nm and its genomic size is around 29.7 kb.[9,10] The SARS-CoV genome contains 14 open reading frames (ORFs) flanked by 5′- and 3′-untranslated regions of 265 and 342 nucleotides, respectively. While all CoVs carry strain-specific accessory proteins encoded by their downstream ORFs, the order of essential genes—the replicase/transcriptase gene, S gene, E gene, M gene, and N gene—is highly conserved.[11] Similar to other known coronaviruses, the SARS-CoV genome expression starts with two long open reading frames (ORFs), ORF1a and ORF1b, which account for two-thirds of the genomic capacity, followed by ORFs encoding S, E, M, and N proteins (Fig. 26.1). In addition to these conserved core genes in coronaviruses, the SARS-CoV

genome contains several accessory genes that are specific to SARS-CoV and their encoded products have no homologue to known proteins. Phylogenetic analysis based on the most conserved gene *ORF1b* indicated that SARS-CoV is distantly related to the group 2 coronaviruses (now the genus *Betacoronavirus*) in the family

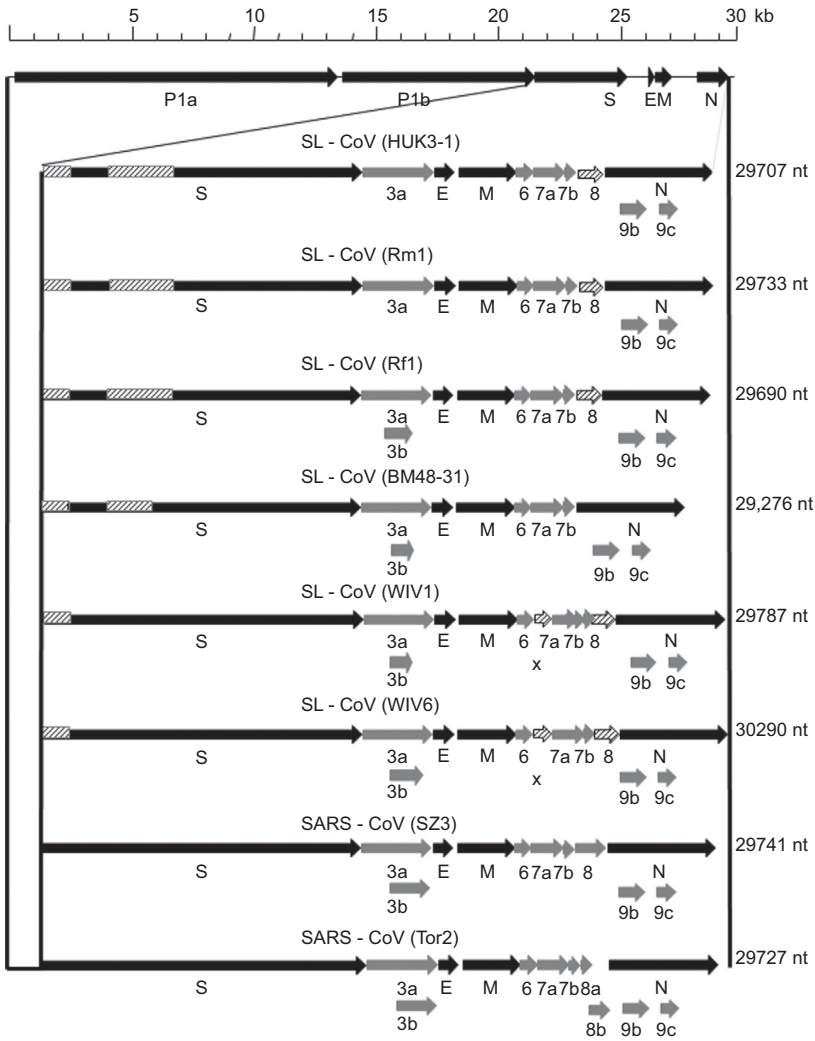

Figure 26.1 Genomic structure of SARS-CoV and bat SL-CoV. The highly conserved genes present in all coronaviruses are shown in *dark-colored arrows* and the betacoronavirus group b-specific ORFs in *light-colored arrows*. The most variable regions are marked with shaded boxes. Rp3, HKU3-1, WIV1, and WIV16 were identified from *R. sinicus* in China; Rm1 and Rf1 from *Rhinolophus macrotis* and *Rhinolophus ferrumequinum*, respectively, in China; BM48-31 from *Rhinolophus blasii*, in Europe; Tor2 from late-phase patient during 2002−2003 SARS outbreak; SZ from civet during 2002−2003 SARS outbreak. * The host of Rp3 was previously identified as *Rhinolophus pearsoni* and later corrected to be *R. sinicus*.[28]

Coronaviridae, and represents a distinct cluster, named group 2b (now the genus *Betacoronavirus* group b; Fig. 26.2).[12,13]

3. The Animal Link

Due to the rapid spread of the disease and the delay in the identification of the causative agent, there was no detailed epidemiological tracing done at the beginning of the

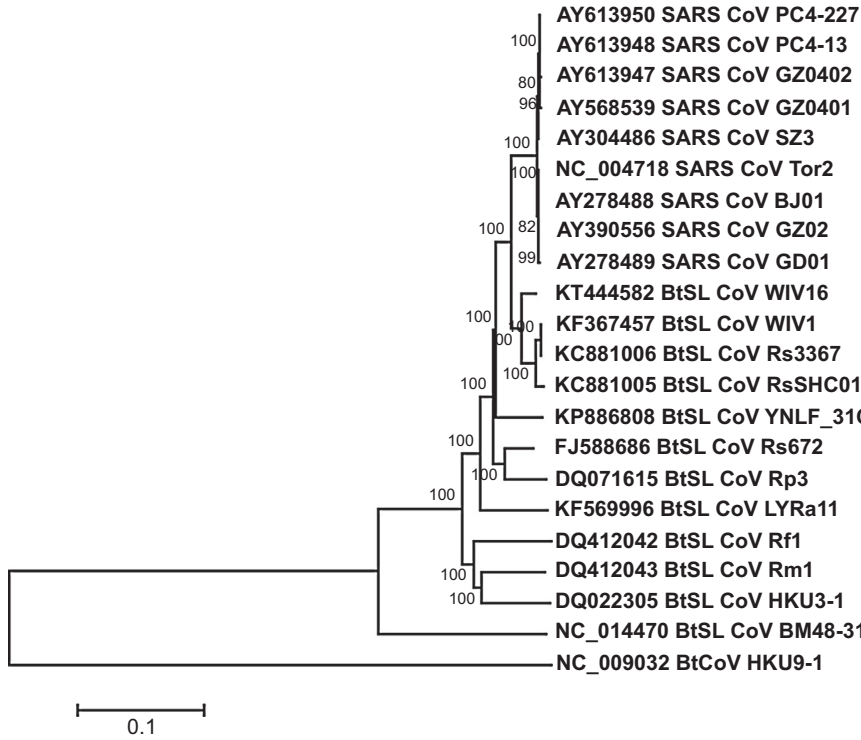

Figure 26.2 Phylogenetic tree of betacoronavirus group b. The phylogenetic tree is generated based on full-length genome sequences of selected SARS-CoVs and bat SL-CoVs using the Neighbor-Joining algorithm in the MEGA4 program[78] with a bootstrap of 1000 replicates. A bat coronavirus BtCoV HKU9 is used as an outgroup.[79] Numbers above branches indicate bootstrap values from 1000 replicates. Scale bar, 0.5 substitutions per site. GD01: SARS-CoV isolate from early-phase patient during 2002−2003 SARS outbreak; Tor2, BJ01: SARS-CoV isolate from late-phase patient during 2002−2003 SARS outbreak; SZ: SARS-CoV isolate from civet during 2002−2003 SARS outbreak; GZ0401/02: SARS-CoV isolate from patient during 2003−2004 SARS outbreak; and PC4-13, PC4-227: SARS-CoV isolate from civet during 2003−2004 SARS outbreak. BtSL-CoV: bat SARS-like CoV. Rp3, HKU3-1, WIV, WIV16, and LYRa11 were identified from *R. sinicus* in China; Rm1 from *Rhinolophus macrotis* in China; Rf1 and YNLC31 C from *Rhinolophus ferrumequinum* in China; and BM48-31 from *Rhinolophus blasii*, in Europe.

outbreaks, and it was therefore impossible to trace the origin of the virus. However, through retrospective investigation, it emerged that the majority of the early index cases were limited in several cities of the Guangdong province and most of them have history of contact directly or indirectly with wildlife animals, including handling, killing, and selling wildlife animals as well as preparing and serving wildlife animal meat in restaurants.[14–16]

As these epidemic regions have a unique dietary tradition favoring freshly slaughtered game meat, there is a huge trafficking and trading industry dedicated to live animal trading in specialized market, the "wet market." Immediately after SARS-CoV was identified as the etiological agent of SARS, studies were conducted in those markets for evidence of SARS-CoV in market animals. One of the earliest and most important studies was conducted by a joint team from Hong Kong and Shenzhen in mainland China.[14] In this investigation, out of 25 samples collected from market animals, SARS-CoV-like viruses were isolated from four out of six masked palm civets (*Paguma larvata*) and one raccoon dog (*Nyctereutes procyonoides*). Antibodies against SARS-CoV were detected in masked palm civets, raccoon dog, and Chinese ferret-badgers (*Melogale moschata*). Genome sequencing indicated that the viruses isolated from civets were almost identical to those from human, suggesting a highly possible zoonotic transmission of SARS-CoV from animal(s) to human.[14] These data indicated that at least three different animal species were infected by a coronavirus that is closely related to SARS-CoV. This important study provided the first direct evidence that SARS-CoV existed in animals, pointing to an animal link of the SARS outbreaks.

Although three animals were identified as susceptible to SARS-CoV infection, the larger sale volume of civets in comparison to other animals in the market made them the target animals of subsequent surveillance studies. The role of civets as a major carrier of SARS-CoV in the markets was further confirmed by serological studies involving much large samples.[17,18]

The most detailed epidemiological data proving a direct civet to human transmission of SARS-CoV was obtained during the investigation of the second wave of SARS outbreaks during December 2003 to January 2004. There were two lines of evidences suggesting a direct transmission. First, all four independent cases had the history of direct or indirect contact with civets. Second, sequencing analysis indicated that sequences derived from human samples were more closely related to those in the civets during that period than those from human samples obtained in the major 2002–2003 outbreaks.[3]

In summary, based on the previously mentioned study findings, it was concluded that the civet to human transmission is a major, if not the only, source of SARS-CoV introduction into the human population.[19–21]

4. Natural Reservoirs of SARS-CoV

Natural reservoir refers to the long-term host of the pathogen of an infectious disease. It is often the case that hosts do not get the disease carried by the pathogen, or the

infection in the reservoir host is subclinical, asymptomatic, and nonlethal. Once discovered, natural reservoirs elucidate the complete life cycle of infectious diseases, which in turn will help to provide effective prevention and control strategies.

As stated earlier, it is clear that civets played a pivotal role in the 2002−2004 outbreaks of SARS in southern China. Culling of civets seemed to be effective in controlling further outbreaks in the region. However, the role of civets as a potential natural reservoir host was less evident and eventually ruled out by several studies. Serological and molecular studies indicated that only civets in the markets were infected with SARS-CoV whereas the populations of civets in the wild or on farms were free of major infections.[18,22,23] Civets produced overt clinical syndromes when experimentally infected with SARS-CoV.[24] Comparative genome sequence analysis indicated that SARS-CoVs in civets experienced rapid mutation, suggesting that the viruses were still adapting to the host rather than persisting in equilibrium expected for viruses in their natural reservoir species.[17,25]

Continuing search for the potential reservoir host of SARS-CoV resulted in the simultaneous discovery of SARS-like coronaviruses (SL-CoVs) in bats by two independent teams in 2005. Using serological and PCR surveillance, both groups discovered that SL-CoVs were present in different horseshoe bats in the genus *Rhinolophus*.[22,26] Complete genome sequence analysis revealed that bat SL-CoVs have an identical genome organization and a nucleotide sequence identity of 88−92% to SARS-CoV (Fig. 26.1; Table 26.2). Except for the S, ORF3, and ORF8 gene products, all deduced aa sequences of the other gene products have a sequence identity above 93% with those of SARS-CoV. The variable regions between SARS-CoV and bat SL-CoV are mainly located in the coding regions for the nonstructural protein 3 (Nsp 3), S protein, ORF3, and ORF8, the products of these genes have aa sequence identity of 87−95%, 76−78%, 82−90%, and 34−80%, respectively. Among the different bat SL-CoVs, the coding regions for these proteins also represent the most variable regions.[27−29]

The phylogenetic analysis indicated that bat SL-CoVs were grouped in the same cluster of SARS-CoV and were only distantly related to other previously known coronaviruses (Fig. 26.2). To date, these bat SL-CoVs represent naturally occurring CoVs that are most closely related to the SARS-CoVs isolated from humans and civets.

Analysis of nonsynonymous and synonymous substitution rates in bat SL-CoVs suggests that these viruses are not experiencing a positive selection pressure that would be expected if horseshoe bats are new host to these viruses. Instead, these data would argue that these viruses have been associated with the bat hosts for a long time.[27,29,30] These observations would support the notion that bats in the genus *Rhinolophus* are the likely natural reservoir hosts of bat SL-CoVs. It can be further postulated that similar bat species may serve as natural reservoirs of viruses with closer evolutionary relationship to the viruses that were responsible for the 2002−2004 SARS outbreaks.

In this context, we and other groups continued the search for the direct progenitor of SARS-CoV and made great progress in the last 10 years following the initial discovery of SL-CoVs in horseshoe bats. First, highly diverse SL-CoVs have been found not only in Chinese but also in European and African bats, indicating a much wider geographic distribution and long evolutionary history of SL-CoVs in different bat populations

Table 26.2 Comparison of Gene Products Between SARS-CoV and Bat SL-CoV

	Gene Product Size (aa)							Amino Acid Sequence Identity With Tor2/sz3 (%)[a]				
Gene/ORF	**Tor2**	**SZ3**	**Rf1**	**Rp3**	**Rm1**	**HKU3-1**	**Rs1**	**Rf1**	**Rp3**	**Rm1**	**HKU3-1**	**Rs672**
P1a	4382	4382	4377	4380	4388	4376	4189	94	96	93	94	94
P1b	2628	2628	2628	2628	2628	2628	2628	98	99	98	98	99
nsp3[b]	1922	1922	1917	1920	1928	1916	1729	92	95	90	92	87
S	1255	1255	1241	1241	1241	1242	1241	76	78	78	78	79
S1	680	680	666	666	666	667	666	63	63	64	6	64
S2	575	575	575	575	575	575	575	92	96	96	94	96
ORF3a	274	274	274	274	274	274	274	86	83	83	82	90
ORF3b	154	154	113	56	56	39	114	89	NA	NA	NA	97
ORF3c	NP	NP	32	NP	NP	NP	NP	NA	NA	NA	NA	NA
E	76	76	76	76	76	76	76	96	100	98	100	100
M	221	221	221	221	221	221	221	97	97	97	99	99
ORF6	63	63	63	63	63	63	63	93	92	92	94	98
ORF7a	122	122	122	122	122	122	122	91	95	93	94	96

ORF7b	44	44	44	44	44	44	44	90	93	93	93	93
ORF8a	39	NP	NP	NP	NP	NP	NP	NA	NA	NA	NA	NA
ORF8b	84	NP	NP	NP	NP	NP	NP	NA	NA	NA	NA	NA
ORF8	NP	122	122	121	121	121	121	80	35	35	34	36
N	422	422	421	421	420	421	422	95	97	97	96	99
ORF9a	98	98	96	97	97	97	98	81	85	90	88	92
ORF9b	70	70	70	70	70	70	70	80	91	91	88	94

NP, not present; *NA*, not applicable.
[a]Tor2 was used for all homology calculations with the exception of ORF8, which is absent in Tor2, the SZ3 was used instead.
[b]The region of nsp3 is highly variable and was calculated alone.

(Table 26.2).[31–37] Second, great genetic diversity of SARS-CoVs were discovered in one particular population of *R. sinicus* in China by a longitudinal surveillance.[31,38] Third and most importantly, two SL-CoV strains were isolated in Vero cells. These two isolates are closely related to the progenitor of the SARS-CoV not only in genomic sequences but also in receptor usage[31,34](Figs. 26.1–26.3; Table 26.1).

5. Molecular Evolution of SARS-CoV in Humans and Animals

Analysis of the large number of SARS-CoV and SL-CoV sequence datasets accumulated since 2004 has clearly demonstrated the importance of virus evolution in cross-species transmission and in pathogenesis. The following is a summary of the major evolutionary findings in host switching, recombination, and virus–receptor interactions.

5.1 Rapid Adaptation of SARS-CoVs in Humans

On the basis of the epidemiological data, the Chinese SARS molecular epidemiology consortium divided the course of the 2002–2004 outbreaks into three stages, the early, middle, and late phases, respectively.[1] The early phase is defined as the period from the first emergence of SARS to the first documented SSE. The middle phase refers to the ensuing events up to the first cluster of SARS cases in a hotel (Hotel M) in Hong Kong, while cases following this cluster fall into the late phase.

Analysis of all the viral sequences available from human patients and animals revealed two major hallmarks of rapid virus evolution during the initial stages of the 2002–2003 outbreaks: (1) All isolates from early patients and market animals contained a 29-nucleotide (nt) sequence in ORF8 that is absent in most of the publicly available human SARS-CoV sequences derived from later phases of the outbreaks; (2) characteristic motif of single-nucleotide variations (SNVs) were identified in SARS-CoVs of different phases and all these SNVs were located in the S gene that codes for the spike protein responsible for attachment to the host cellular receptor.[25] All SARS-CoV isolates from epidemic countries and regions outside mainland China could be traced to Guangdong or Hong Kong based on the S-gene SNV motif.[23,39]

During the second sporadic outbreaks of 2003–2004, it was shown that the SARS-CoV sequences from index patients were almost identical to that from civets collected in the same period and all retained the 29-nt sequence in the ORF8 gene. The mild disease symptoms associated with these viruses and the lack of rapid human-to-human transmission provided further evidence that the rapid adaptation of the SARS-CoV in the first major outbreak of 2002–2003 was essential for its establishment and pathogenesis in humans.

With the available genomic variation data and the sampling time, it is now possible to calculate the neutral mutation rate and to estimate the date for the most recent common ancestors (MRCAs) of SARS-CoV. The estimate obtained is around 8.00×10^{-6} nt^{-1} day^{-1}, suggesting that SARS-CoV evolves at a relatively constant

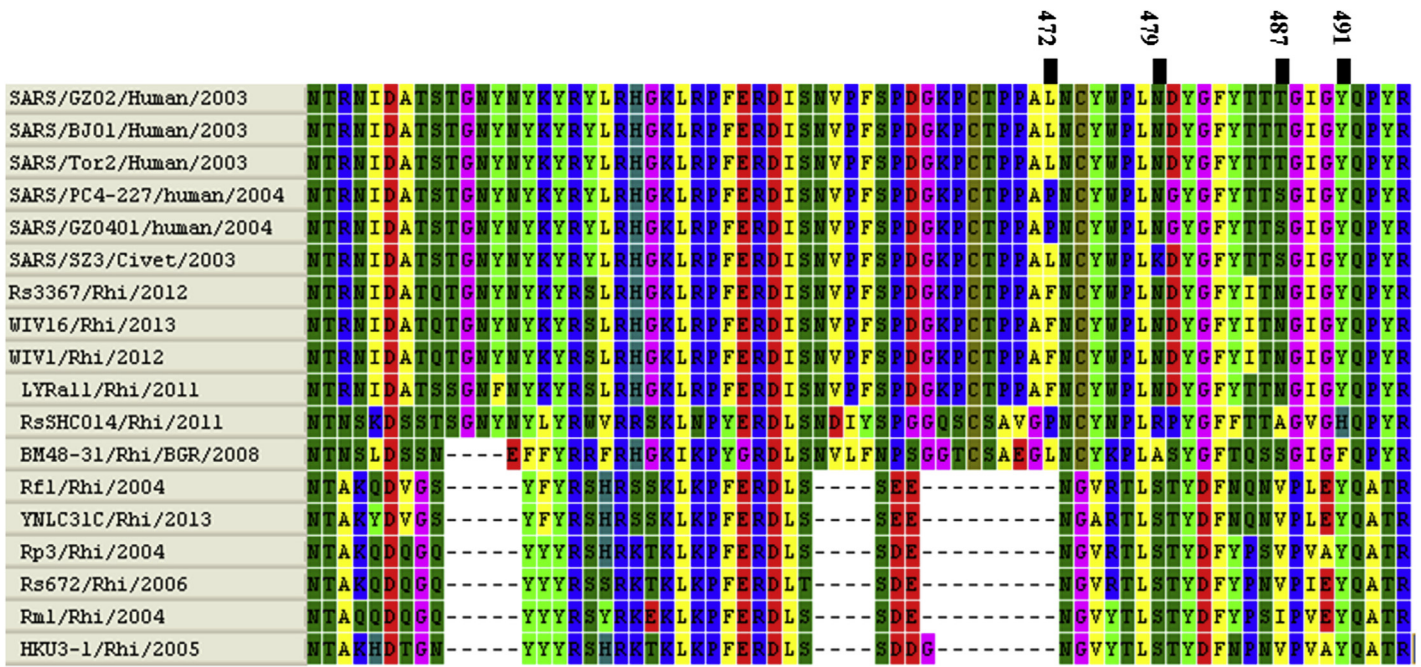

Figure 26.3 Alignment of amino acid sequences covering the receptor-binding motif from viruses of different species origin. GD01: SARS-CoV isolate from early-phase patient during 2002−2003 SARS outbreak; Tor2, BJ01: SARS-CoV isolate from late-phase patient during 2002−2003 SARS outbreak; SZ: SARS-CoV isolate from civet during 2002−2003 SARS outbreak; GZ0402: SARS-CoV isolate from patient during 2003−2004 SARS outbreak; and PC4-227: SARS-CoV isolate from civet during 2003−2004 SARS outbreak. * indicates the two key residues 479 and 487. Rp3, HKU3-1, WIV, WIV16, and LYRa11 were identified from *R. sinicus* in China; Rm1 from *Rhinolophus macrotis* in China; Rf1 and YNLC31C from *Rhinolophus ferrumequinum* in China; and BM48-31 from *Rhinolophus blasii*, in Europe.

neutral rate both in humans and palm civet. From these calculations, it was estimated that the MRCAs for palm civet and humans of different transmission lineages lie in mid-November 2002. This estimate was consistent with the first observed SARS case around November 16, 2002 in Foshan, Guangdong.[1,2,25]

5.2 Generation of Viral Genetic Diversity by Recombination

At the present time, at least 33 full-length genome sequences of bat SL-CoVs were determined.[22,26–29,32,34–37,40] Shown in Fig. 26.1 is a comparison of the genome structures for seven selected bat SL-CoVs and one each of civet and human SARS-CoV isolates. All bat SL-CoVs, with the exception of HKU3-8[29] and BM48-31,[32] contain the 29-nt sequence in ORF8, which is present in SARS-CoV from early-phase patients and civets, indicating the common ancestor between civet SARS-CoV and bat SL-CoV. The SL-CoV HKU3-8 contained a 26-nt deletion that is located 14 nt downstream from the commonly observed 29-nt deletion, and the BM48-31 completely lost the ORF8, indicating that the ORF8 coding region is a "hotspot" for deletions.

SL-CoVs from different bat species share 88–97% nt identity among themselves, indicating that the genetic diversity of SL-CoVs in bats is much greater than that observed among civet or human isolates. The most dramatic sequence difference between human SARS-CoV and bat SL-CoV is in the S protein that has 76–97% aa identity for the whole S protein and 64–95% aa identity for the N-terminal region (or the S1 region; Table 26.2). This great genetic diversity observed among bat SL-CoVs and the major difference between the S1 regions of SL-CoV and SARS-CoV S proteins clearly demonstrated that bats are natural reservoirs of human SARS-CoV.

It is well documented that the positive-sense ssRNA genomes of coronaviruses are prone to homologous recombination during coinfection of different coronaviruses and that recombination plays an important role in generating new coronavirus species, in facilitating cross-species transmission and in modulating virus virulence.

Several studies provided evidence for coinfection and recombination came from analysis of SL-CoVs in bats.[29,41–44] It was further revealed that recombination can occur at multiple sites along the SL-CoV genome.[11,28,29,31,34,41] For example, detailed sequence analysis of two genotypes of bat SL-CoV, Rp3 and Rs672 (both were identified from *R. sinicus*), suggested that they may represent a recombinant of two bat SL-CoVs and one of them is more closely related to the human SARS-CoVs.[28,41] During 2015 and 2016, two teams reported a full-length ORF8 that shares higher sequence similarities to the SARS-CoV GZ02 and civet SARS-CoV SZ3 than previously detected SL-CoVs.[37,40] These results suggest that SARS-CoV most likely originated from different bat SL-CoVs via a complicated evolutionary path that involved recombination events.

5.3 Receptor Usage and Evolutionary Selection

The S protein of coronavirus is responsible for attachment to cellular receptor to initiate the first step of virus infection. The angiotensin-converting enzyme 2 (ACE2) was identified as a main functional receptor for SARS-CoV.[45] Further analysis demonstrated that the region covering aa 318–520 of S protein is the key receptor-

binding domain (RBD), which is both essential and sufficient to bind the human ACE2 molecule in vitro.[46] Detailed analysis of the crystal structure of the RBD−ACE2 complex revealed that 19 key residues have close contact with the receptor molecule, which are located from aa 424 to 474. This region is termed the receptor-binding motif (RBM).[47]

When the existing epidemiological data was analyzed in combination with the data on infectivity of SARS-CoV isolated in humans at the different phases of the outbreaks and SARS-CoV isolates in civets, a clear correlation could be established between the evolution of the S proteins and virus infectivity. It was observed that the S protein is the fastest evolving protein of SARS-CoV during interspecies transmission from animal to human and in the following phases of human to human transmission. The majority of the mutations are located in the S1 domain (31 of a total of 48 SNVs), particularly in the RBD.[1,46] The interaction analysis between the S proteins of different isolates and the ACE2 molecules demonstrated that two aa residues in the S protein, aa 479 and aa 487, played a key role in virus infectivity.[48,49] For aa residue 479, all 2002−2003 human isolates contain asparagine (N). The palm civet isolates seem to have variable aa residues at this position, all 2002−2003 and some 2003−2004 civet isolates have lysine (K) while other 2003−2004 isolates have either asparagine (N) or arginine (R). For aa residue 487, all isolates including those from early- and middle-phase patients, civets of 2002−2003 and 2003−2004, have a codon for serine (S), whereas all isolates from 2002−2003 late-phase human patients have a codon for threonine (T) (Fig. 26.3). When examined using an HIV-based pseudovirus infection assay, S proteins with all combinations of residues 487/479 could efficiently use the civet ACE2 as an entry receptor, but showed different infectivity in human ACE2-mediated infection.[48,49] The combination of N479/T487 had the highest infectivity, N479/S487 medium infectivity, and K479/S487 the lowest, which almost abolished the infection. These results demonstrated elegantly at the molecular interface that the rapid evolution of the S protein, especially at the aa positions important for host receptor engagement, was essential for the adaptation to and establishment of an effective and productive human infection.

When the genome sequences of SL-CoVs were analyzed, it became evident that the N-terminal regions of their S proteins are the most divergent among themselves, as well as with the SARS-CoV. As shown in Fig. 26.3, bat SL-CoVs can be grouped into three groups based on the RBM sequences. The strains discovered early are close to each other and have a major sequence difference involving deletions of 17−18 aa right in the middle of RBM. We have since demonstrated experimentally that SL-CoV S proteins are unable to use ACE2 molecule, regardless of its origin, as a functional receptor. The second group, identified from European bats, has deletions of 4 aa.[32] The third group, discovered recently, has no deletion and contains an identical size as the SARS-CoV in the S protein (Fig. 26.3).[31,34,35] As predicted from their S sequences, three isolates from the third group, SL-CoV—WIV1, WIV16, and SHC014, have been shown to be able to use ACE2 for cellular entry, even though these S proteins still have slight difference at the key aa involved in direct interaction with ACE2.[31,34,50] Most importantly, the SHC014 can replicate well in transgenic mice containing human ACE2, and it caused tissue damage in tested animals.[50]

6. Coronavirus Surveillance in Wildlife Animals

Zoonosis contributes to the majority of emerging disease in the last 30 years, many of them originated from wildlife animals.[51−55] The story of SARS is just one of such examples that spectacularly demonstrated the seamless evolution of a bat virus into a human pathogen responsible for one of the most severe global pandemic outbreaks in modern history of mankind. In general, pathogens carried by wildlife reservoir animals usually do not cause clinical symptoms and they lie dormant until they spill over into and cause diseases in domestic animals or humans. Classical outbreak response measures, such as those deployed during the SARS outbreaks, are still useful, but no longer sufficient for early detection and prevention of major infectious disease outbreaks in the 21st century.

With the demonstration of an increasing number of spillover events that led to severe disease outbreaks in human and domestic animals, we believe it is paramount that from now on we include active surveillance of wildlife animals as part of an integrated infectious disease prevention and control strategy. Surveillance of wildlife animals has also been made more feasible and productive, thanks to the advance in modern molecular techniques including PCR with virus group-specific primers, virus discovery using next generation high-throughput sequencing technologies, and high density virus microarrays.[56−63] Since the SARS outbreaks, especially after the discovery of SL-CoVs in bats, there is a significant surge in international effort for surveillance of coronaviruses in wildlife animals. Before the SARS outbreak, there were only 10 coronaviruses with complete genomes sequenced. This number has increased more than sixfold as a result of the active surveillance works conducted around the world.[27−29,31,32,34,40,55,64−73] Although this only marks the beginning of our understanding of coronaviruses in wildlife animals, it is fair to say that we have learnt a lot more about coronaviruses in general than the past 50 years or so; during that period studying of viruses was only possible and called for in response to disease outbreaks. Based on phylogenetic analysis of the large number of bat coronavirus sequences available presently, it is postulated that all known disease-causing coronaviruses previously identified in humans or animals originated from bat strains.[31,34,43,55] This hypothesis was unfortunately proved by the outbreak of another SARS-like disease, Middle East Respiratory Syndrome (MERS), which was caused by a novel coronavirus (previously named HCoV-EMC, now MERS-CoV) and supposed to originate from bats.[74] Even though the MERS-CoV-like viruses found in bats are not the direct progenitor of the MERS-CoV, the highly genetic diversity of these bat viruses is likely the gene sources for the deadly pathogen in humans, just like that for SARS-CoV.[71,72,75−77]

7. Concluding Remarks

The emergence of SARS-CoV has had a huge impact on the global health and economy. It served as a warning to what may come out of a seemingly harmless virus−reservoir equilibrium in bats or any other wildlife species. At the same time, the experience

gained from the SARS outbreaks and the following in-depth studies on SARS-like coronaviruses has provided and will continue to provide invaluable knowledge and guideline to our future fight against new and emerging infectious diseases. One of the major lessons is that we need to pay much more attention to the reservoir species in understanding the genetic diversity of different viruses, the intricate interplay at the virus−host interface, and the major factors responsible for the disturbance of virus−host equilibrium, which in turn trigger spillover events leading to disease outbreaks.

References

1. Chinese SARS Molecular Epidemiology Consortium. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* 2004;**303**(5664):1666−9.
2. Zhao GP. SARS molecular epidemiology: a Chinese fairy tale of controlling an emerging zoonotic disease in the genomics era. *Philos Trans R Soc Lond B Biol Sci* 2007;**362**(1482):1063−81.
3. Wang M, Yan M, Xu H, Liang W, Kan B, Zheng B, et al. SARS-CoV infection in a restaurant from palm civet. *Emerg Infect Dis* 2005;**11**(12):1860−5.
4. Lim W, Ng KC, Tsang DN. Laboratory containment of SARS virus. *Ann Acad Med Singap* 2006;**35**(5):354−60.
5. WHO. Severe acute respiratory syndrome (SARS). *Wkly Epidemiol Rec* 2003;**78**:81−8.
6. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 2003;**348**(20):1953−66.
7. Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 2003;**361**(9366):1319−25.
8. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 2003;**348**(20):1967−76.
9. Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, et al. The genome sequence of the sars-associated coronavirus. *Science* 2003;**300**(5624):1399−404.
10. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003;**300**(5624):1394−9.
11. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol* 2010;**84**(7):3134−46.
12. Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LLM, et al. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* 2003;**331**(5):991−1004.
13. de Groot R, Baker S, Baric R, Enjuanes L, Gorbalenya A, Holmes K, et al. Family Coronaviridae. In: King A, Adams M, Cartens E, Lefkowitz E, editors. *Virus taxonomy; ninth report of the international committee on taxonomy of viruses*. San Diego, CA: Academic Press; 2012. p. 806−28.
14. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 2003;**302**(5643):276−8.

15. Xu HF, Wang M, Zhang ZB, Zou XZ, Gao Y, Liu XN, 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* 2004;**38**(2):81−3.
16. Xu RH, He JF, Evans MR, Peng GW, Field HE, Yu DW, et al. Epidemiologic clues to sars origin in China. *Emerg Infect Dis* 2004;**10**(6):1030−7.
17. Kan B, Wang M, Jing H, Xu H, Jiang X, Yan M, 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* 2005;**79**(18):11892−900.
18. Tu C, Crameri G, Kong X, Chen J, Sun Y, Yu M, et al. Antibodies to SARS coronavirus in civets. *Emerg Infect Dis* 2004;**10**(12):2244−8.
19. Wang LF, Shi Z, Zhang S, Field H, Daszak P, Eaton BT. Review of bats and SARS. *Emerg Infect Dis* 2006;**12**(12):1834−40.
20. Wang LF, Eaton BT. Bats, civets and the emergence of SARS. *Curr Top Microbiol Immunol* 2007;**315**:325−44.
21. Shi Z, Hu Z. A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res* 2008;**133**(1):74−87.
22. Poon LL, Chu DK, Chan KH, Wong OK, Ellis TM, Leung YH, et al. Identification of a novel coronavirus in bats. *J Virol* 2005;**79**(4):2001−9.
23. Lan YC, Liu TT, Yang JY, Lee CM, Chen YJ, Chan YJ, et al. Molecular epidemiology of severe acute respiratory syndrome-associated coronavirus infections in Taiwan. *J Infect Dis* 2005;**191**(9):1478−89.
24. Wu DL, Tu CC, Xin C, Xuan H, Meng QW, Liu YG, et al. Civets are equally susceptible to experimental infection by two different severe acute respiratory syndrome coronavirus isolates. *J Virol* 2005;**79**(4):2620−5.
25. Song HD, Tu CC, Zhang GW, Wang SY, Zheng K, Lei LC, et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci USA* 2005;**102**(7):2430−5.
26. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci USA* 2005; **102**(39):14040−5.
27. Ren W, Li W, Yu M, Hao P, Zhang Y, Zhou P, et al. Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. *J Gen Virol* 2006;**87**(Pt 11):3355−9.
28. Zhang Y, Zhang H, Dong X, Yuan J, Yang X, Zhou P, et al. Hantavirus outbreak associated with laboratory rats in Yunnan, China. *Infect Genet Evol* 2010;**10**(5):638−44.
29. Lau SK, Li KS, Huang Y, Shek CT, Tse H, Wang M, 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* 2010;**84**(6):2808−19.
30. Tang X, Li G, Vasilakis N, Zhang Y, Shi Z, Zhong Y, et al. Differential stepwise evolution of SARS coronavirus functional proteins in different host species. *BMC Evol Biol* 2009;**9**:52.
31. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 2013; **503**(7477):535−8.
32. Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, Goettsche M, 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* 2010;**84**(21):11336−49.

33. Tong S, Conrardy C, Ruone S, Kuzmin IV, Guo X, Tao Y, et al. Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg Infect Dis* 2009;**15**(3):482−5.
34. Yang XL, Hu B, Wang B, Wang MN, Zhang Q, Zhang W, et al. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of SARS coronavirus. *J Virol* 2015;**90**(6):3253−6.
35. He B, Zhang Y, Xu L, Yang W, Yang F, Feng Y, et al. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. *J Virol* 2014;**88**(12):7070−82.
36. Yang L, Wu Z, Ren X, Yang F, He G, Zhang J, et al. Novel SARS-like betacoronaviruses in bats, China, 2011. *Emerg Infect Dis* 2013;**19**(6):989−91.
37. Wu Z, Yang L, Ren X, Zhang J, Yang F, Zhang S, et al. ORF8-Related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J Infect Dis* 2016;**213**(4):579−83.
38. Wang M-N, Zhang W, G Y-T, H B, G X-Y, Y X-L, et al. Longitudinal surveillance of SARS-like coronaviruses in bats by quantitative real-time PCR. *Virol Sin* 2016;**31**(1): 78−80.
39. Tang JW, Cheung JL, Chu IM, Ip M, Hui M, Peiris M, et al. Characterizing 56 complete SARS-CoV S-gene sequences from Hong Kong. *J Clin Virol* 2007;**38**(1):19−26.
40. Lau SK, Feng Y, Chen H, Luk HK, Yang WH, Li KS, et al. Severe acute respiratory syndrome (sars) coronavirus ORF8 protein is acquired from sars-related coronavirus from greater horseshoe bats through recombination. *J Virol* 2015;**89**(20):10532−47.
41. Hon CC, Lam TY, Shi ZL, Drummond AJ, Yip CW, Zeng F, 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* 2008;**82**(4):1819−26.
42. Cui J, Han N, Streicker D, Li G, Tang X, Shi Z, et al. Evolutionary relationships between bat coronaviruses and their hosts. *Emerg Infect Dis* 2007;**13**(10):1526−32.
43. Vijaykrishna D, Smith GJ, Zhang JX, Peiris JS, Chen H, Guan Y. Evolutionary insights into the ecology of coronaviruses. *J Virol* 2007;**81**(8):4012−20.
44. Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, Li LF, et al. Prevalence and genetic diversity of coronaviruses in bats from China. *J Virol* 2006;**80**(15):7481−90.
45. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003;**426**(6965): 450−4.
46. Wong SK, Li W, Moore MJ, Choe H, Farzan M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* 2004; **279**(5):3197−201.
47. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005;**309**(5742):1864−8.
48. Qu XX, Hao P, Song XJ, Jiang SM, Liu YX, Wang PG, 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* 2005;**280**(33):29588−95.
49. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* 2005;**24**(8):1634−43.
50. Menachery VD, Yount Jr BL, Debbink K, Agnihothram S, Gralinski LE, Plante JA, et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* 2015;**21**(12):1508−13.
51. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature* 2008;**451**(7181):990−3.

52. Chomel BB, Belotto A, Meslin FX. Wildlife, exotic pets, and emerging zoonoses. *Emerg Infect Dis* 2007;**13**(1):6−11.
53. Bengis RG, Leighton FA, Fischer JR, Artois M, Morner T, Tate CM. The role of wildlife in emerging and re-emerging zoonoses. *Rev Sci Tech* 2004;**23**(2):497−511.
54. Woolhouse ME, Gowtage-Sequeria S. Host range and emerging and reemerging pathogens. *Emerg Infect Dis* 2005;**11**(12):1842−7.
55. Liang YZ, Wu LJ, Zhang Q, Zhou P, Wang MN, Yang XL, et al. Cloning, expression, and antiviral activity of interferon beta from the Chinese microbat. *Myotis Davidii Virol Sin* 2015;**30**(6):425−32.
56. Ng TF, Manire C, Borrowman K, Langer T, Ehrhart L, Breitbart M. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *J Virol* 2009;**83**(6):2500−9.
57. Yanai-Balser GM, Duncan GA, Eudy JD, Wang D, Li X, Agarkova IV, et al. Microarray analysis of *Paramecium bursaria* chlorella virus 1 transcription. *J Virol* 2010;**84**(1): 532−42.
58. Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, et al. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* 2003;**185**(20): 6220−3.
59. Gaynor AM, Nissen MD, Whiley DM, Mackay IM, Lambert SB, Wu G, et al. Identification of a novel polyomavirus from patients with acute respiratory tract infections. *PLoS Pathog* 2007;**3**(5):e64.
60. Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, et al. Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* 2012;**86**(8):4620−30.
61. Lipkin WI, Anthony SJ. Virus hunting. *Virology* 2015;**479**−**480**:194−9.
62. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *Mbio* 2015;**6**(5):e01491−515.
63. Barzon L, Lavezzo E, Militello V, Toppo S, Palu G. Applications of next-generation sequencing technologies to diagnostic virology. *Inter J Mol Sci* 2011;**12**(11):7861−84.
64. Vijgen L, Keyaerts E, Lemey P, Maes P, Van Reeth K, Nauwynck H, et al. Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. *J Virol* 2006;**80**(14): 7270−4.
65. Vijgen L, Keyaerts E, Moes E, Thoelen I, Wollants E, Lemey P, et al. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J Virol* 2005;**79**(3):1595−604.
66. Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol* 2005;**79**(2):884−95.
67. Woo PC, Lau SK, Huang Y, Yuen KY. Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med (Maywood)* 2009;**234**(10):1117−27.
68. Woo PC, Lau SK, Lam CS, Lai KK, Huang Y, Lee P, et al. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. *J Virol* 2009;**83**(2):908−17.
69. Woo PC, Lau SK, Yip CC, Huang Y, Tsoi HW, Chan KH, et al. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. *J Virol* 2006;**80**(14):7136−45.

70. Alekseev KP, Vlasova AN, Jung K, Hasoksuz M, Zhang X, Halpin R, et al. Bovine-like coronaviruses isolated from four species of captive wild ruminants are homologous to bovine coronaviruses, based on complete genomic sequences. *J Virol* 2008;**82**(24): 12422−31.
71. Corman VM, Baldwin HJ, Tateno AF, Zerbinati RM, Annan A, Owusu M, et al. Evidence for an ancestral association of human coronavirus 229E with bats. *J Virol* 2015;**89**(23): 11858−70.
72. Hu B, Ge X, Wang LF, Shi Z. Bat origin of human coronaviruses. *Virol J* 2015;**12**(1):221.
73. Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, 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* 2012;**86**(7): 3995−4008.
74. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New Engl J Med* 2012; **367**(19):1814−20.
75. Yang Y, Liu C, Du L, Jiang S, Shi Z, Baric RS, et al. Two mutations were critical for bat-to-human transmission of Middle East respiratory syndrome coronavirus. *J Virol* 2015;**89**(17): 9119−23.
76. Wang Q, Qi J, Yuan Y, Xuan Y, Han P, Wan Y, et al. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 2014;**16**(3): 328−37.
77. Lau SK, Li KS, Tsang AK, Lam CS, Ahmed S, Chen H, 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* 2013;**87**(15): 8638−50.
78. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;**24**(8):1596−9.
79. Lau SK, Poon RW, Wong BH, Wang M, Huang Y, Xu H, et al. Coexistence of different genotypes in the same bat and serological characterization of Rousettus bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. *J Virol* 2010;**84**(21):11385−94.