SHORT REPORT

Complete genome analysis of a SARS-like bat coronavirus identified in the Republic of Korea

Yongkwan Kim[1,2] · Kidong Son[1] · Young-Sik Kim[2] · Sook-Young Lee[2] · Weonhwa Jheong[1] · Jae-Ku Oem[2]

Abstract

Bats have been widely known as natural reservoir hosts of zoonotic diseases, such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) caused by coronaviruses (CoVs). In the present study, we investigated the whole genomic sequence of a SARS-like bat CoV (16BO133) and found it to be 29,075 nt in length with a 40.9% G+C content. Phylogenetic analysis using amino acid sequences of the ORF 1ab and the spike gene showed that the bat coronavirus strain 16BO133 was grouped with the Beta-CoV lineage B and was closely related to the JTMC15 strain isolated from *Rhinolophus ferrumequinum* in China. However, 16BO133 was distinctly located in the phylogenetic topology of the human SARS CoV strain (Tor2). Interestingly, 16BO133 showed complete elimination of ORF8 regions induced by a frame shift of the stop codon in ORF7b. The lowest amino acid identity of 16BO133 was identified at the spike region among various ORFs. The spike region of 16BO133 showed 84.7% and 75.2% amino acid identity with Rf1 (SARS-like bat CoV) and Tor2 (human SARS CoV), respectively. In addition, the S gene of 16BO133 was found to contain the amino acid substitution of two critical residues (N479S and T487 V) associated with human infection. In conclusion, we firstly carried out whole genome characterization of the SARS-like bat coronavirus discovered in the Republic of Korea; however, it presumably has no human infectivity. However, continuous surveillance and genomic characterization of coronaviruses from bats are necessary due to potential risks of human infection induced by genetic mutation.

Keywords Bat · SARS-like coronavirus · Zoonotic disease · Frame shift · Whole genome

Introduction

Coronaviruses (CoVs) are enveloped viruses containing a single-stranded, positive-sense RNA genome of approximately 27–32 kb [1]. Currently, CoVs are grouped under four distinct genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus* [2, 3].

Bat species have been recognized as major reservoirs of several emerging infectious diseases, such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) [4–6]. SARS is caused by a member of the *Betacoronavirus* genus and is the first global pandemic disease that has emerged in the Guangdong Province of China in 2002. SARS has spread to 25 countries across five continents, infecting 8096 people worldwide with a 9.5% (774/8096) fatality [7–9].

The four structural proteins (S, E, M, and N) are essential for viral entry and assembly. The S gene is the most important structural protein. The receptor-binding motif (RBM) within the receptor-binding domain (RBD) located in the S gene determines host tropism by binding angiotensin-converting enzyme 2 (ACE2) receptor [10, 11]. The RBD has two critical residues (N479 and T487) that play key roles in ACE2 receptor recognition and binding associated with human transmission [7, 12].

✉ Jae-Ku Oem
jku0623@jbnu.ac.kr

1 Environmental Health Research Department, National Institute of Environmental Research, Hwangyeong-ro 42, Seo-gu, Incheon, Republic of Korea

2 Department of Veterinary Infectious Diseases, College of Veterinary Medicine, Chonbuk National University, Jeonju, Republic of Korea

Novel coronaviruses are continuously being discovered in bat species around the world, especially in China [7, 13]. Due to relatively close geographic locations of bat species between China and the Republic of Korea, the surveillance of CoV prevalence and the analysis of their genetic information may be crucial for preventing a future outbreak [14]. However, there have been few investigations into SARS-related bat Beta-CoV prevalence [4]. In addition, whole genome analysis of SARS-related bat Beta-CoV has not yet been carried out in the Republic of Korea.

Together with the fact that bats are reservoirs of CoVs, genetic information about these CoVs may provide valuable information regarding the possible risk of these viruses infecting humans. In the present study, the complete genome sequence of SARS-related Beta-CoV (16BO133) isolated from *Rhinolophus ferrumequinum* was first characterized. The genome of 16BO133 was then compared with that of reference CoVs to demonstrate genetic diversity and a potential genetic feature associated with host tropism.

Results and discussion

An oral swab was collected from bats living in their natural habitat in 2016. Bats were captured using a net for collection of oral swabs and were released immediately after sampling. Oral swab samples were kept in a viral transport medium at 4 °C. The oral swab sample was suspended in 1% antibiotic–antimycotic solution (Corning, USA) diluted in phosphate-buffered saline (PBS), and clarified by centrifugation at 3500×*g* for 10 min. RNA from the 200 μL sample was extracted with the QIAamp® Viral RNA mini kit (Qiagen, Germany) and eluted in 60 μL RNase-free water. cDNA was synthesized using a PrimeScript First Strand cDNA Synthesis kit (Takara, Japan) according to the manufacturer's instructions. Bat-CoV screening was performed by a pancoronavirus PCR method based on primers as follows: (Corona forward, 5′-GGTTGGGACTATCCTAAG TGTGA-3′ and Corona reverse, 5′-CCATCATCAGATAG AATCATCATA-3′). The pancoronavirus primers were used to amplify and sequence a 440-bp segment of the highly conserved RNA-dependent RNA polymerase (RdRp) gene. Fifty-nine pairs of primers were synthesized by the Genotech corporation (Daejeon, Korea) and PCR was performed using an ABI 9800 GeneAmp system (Applied Biosystems, Foster City, CA, USA). The products were purified using a QIAquick gel extraction kit (Qiagen, Germany) according to the manufacturer's instructions. The purified PCR products were sequenced using the BigDye® Terminator Cycle Sequencing kit version 1.1 (Applied Biosystems, Foster City, CA, USA) and an ABI 3730 DNA sequencer (Applied Biosystems, Foster City, CA, USA). Whole genome sequences were submitted to GenBank (accession number KY938558).

The nucleotide and amino acid sequences were aligned and compared to CoV sequences available from the GenBank database using ClustalW software implemented in BioEdit version 7.0.9.0. The phylogenetic trees were drawn using the neighboring joining method using the maximum composite likelihood model with MEGA 7 software. The bootstrap values were calculated with 1000 replicates.

The amino acid sequences of ORF 1ab and spike gene were analyzed for phylogenetic characterization. 16BO133 was grouped with the SARS-related Beta-CoV lineage B, not only due to sequence similarity with ORF 1ab but also with the spike gene (Fig. 1). The RF 1ab and spike amino acids were closely related to JTMC15. However, 16BO133 was distinctly located in the phylogenetic topology of the human SARS CoV strain (Tor2, Urbani, Frankfurt1, and ShanghaiQXC1).

The whole genomic sequence of 16BO133 was 29,075 nt in length with G+C contents of 40.9%. As shown in Table 1, 16BO133 has a similar genome organization to other SARS-related Beta-CoVs, such as JTMC15, Rf1, and Tor2. The 16BO133 showed a high amino acid identity ranging from 93.8% to 100% with JTMC15. However, it showed considerably lower nucleotide identity ranging from 75.2 to 99.5% with Rf1 and Tor2 (Table 1). In addition, a complete deletion of amino acids was observed in the ORF8 region, which is similar to JTMC15 (Table 1). The spike gene nucleotides of 16BO133 showed extensive variations compared to other SARS-related bat Beta-CoV (Rf1) and human SARS CoV (Tor2), thereby resulting in a low amino acid identity. Amino acid identities of 16BO133 spike region with Rf1 and Tor2 were 84.7% and 75.2%, respectively (Table 1).

As shown in Supplementary Fig. 1, the RBM (aa 426–518) located in the S protein showed 18 amino acid deletions (aa 433–437 and 457–469) including critical residues, N479S and T487V. The regions corresponding to TGNYN (433–437) and NVPFSPDGKPCTP (457–469) in human SARS CoV (Tor2) were identified as the major deletion sites in 16BO133. In addition, the insertion of the two nucleotides (cytosine and threonine) was observed in front of the stop codon of ORF7b in 16BO133 (Supplementary Fig. 2). This feature induces a frame shift of the stop codon, resulting in the complete elimination of ORF8.

The bats discovered in the Republic of Korea are considered to be insectivores, and 23 species were reported to exist in this region in a previous study [4]. Recently, wildlife and human contact has increased due to the rapid urbanization. People think that bats are not dangerous because they either living in caves or in abandoned mines. In the present study, SARS-related bat Beta-CoV was identified from *R. ferrumequinum* in an abandoned mine at the Jeonbuk province. Recently, some people visited the abandoned mine out of curiosity, not realizing the risk of exposure to CoV infections upon contact with bat
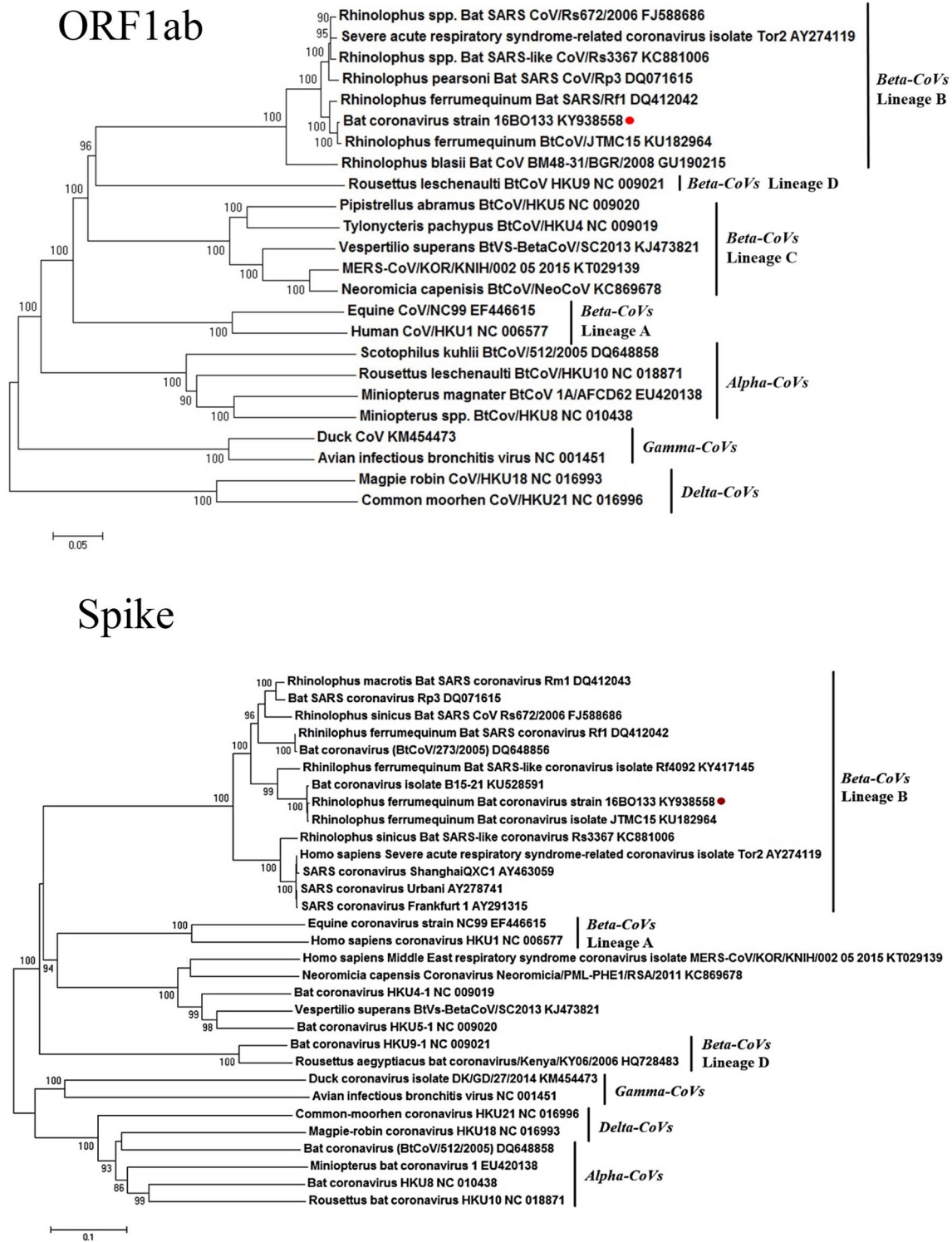
ORF1ab

Spike

Fig. 1 Phylogenetic analysis using whole genome sequences of ORF 1ab with reference strains. The phylogenetic trees were drawn using the neighboring joining method using the maximum composite likelihood model with MEGA 7 software. The bootstrap values were calculated with 1000 replicates. Phylogenetic analysis using whole genome sequences of the spike region with reference strains. The phylogenetic trees were drawn using the neighboring joining method using the maximum composite likelihood model with MEGA 7 software. The bootstrap values were calculated with 1000 replicates

Table 1 Comparison of ORF amino acid identities of 16BO133 and other SARS-CoVs

ORF	16BO133	JTMC15		Rf1		Tor2	
	Length	Length	% Identity	Length	% Identity	Length	% Identity
1a	4184	4184	99.3	4377	98.2	4377	93.4
1b	2646	2646	99.5	2628	98.2	2641	98.1
S	1236	1236	99.4	1241	84.7	1255	75.2
3a	270	274	98.5	274	97.0	274 (ORF3)[a]	85.2
3b	114	114	99.1	114	96.5	154 (ORF4)[a]	89.5
E	76	76	97.4	76	97.4	76	98.7
M	221	221	99.5	221	99.5	221	97.3
6	63	63	98.4	63	98.4	63 (ORF7)[a]	92.1
7a	122	122	97.5	122	95.9	122 (ORF8)[a]	89.3
7b	48	52	93.75	44	97.7	44 (ORF9)[a]	93.2
8	–	–	–	122	–	39 (ORF10)[a]	–
N	420	420	99.3	421	97.1	422	94.3
9a	97	97	95.9	97	94.8	98 (ORF13)[a]	75.3
9b	70	70	100	70	94.3	70 (ORF14)[a]	84.3

Abbreviation and accession numbers: JTMC15, KU182964; Rf1, DQ412042; Tor2, AY274119

[a]ORF 3a, 3b, 6, 7a, 7b, 8, 9a, and 9b are described as ORF 3, 4, 7, 8, 9, 10, 13, and 14 in Tor2

carriers. Therefore, people should keep in mind that bats can spread diseases to humans and should refrain from visiting abandoned mines.

The S gene associated with the spike protein is divided into S1 and S2 domains [15–17]. The S gene is composed of distinct N-terminal (S1) and conserved C-terminal (S2) domains. The S1 domain is prone to have high mutation rates as the virus evolves because it is the major antigenic factor. Therefore, it is thought to be the main reason that the spike protein of 16BO133 has the lowest amino acid identity (75.2%) compared to human SARS CoV (Tor2) within various ORFs.

The S1 domain contains a receptor-binding domain (RBD), which mediates receptor binding of the virus to host cells and determines the host spectrum. The RBM (aa 426 to 518) within the RBD (aa 319 to 518) is the most important motif for recognizing the host receptor, human angiotensin-converting enzyme 2 (ACE2), and it is a major antigenic determinant required to elicit the production of neutralizing antibodies. The RBM has two critical residues, N479 and T487, which play key roles in receptor recognition and binding [15]. The substitution of these two critical residues can completely eliminate viral binding to the human ACE2 receptor [12]. However, substitution of ether residue alone has no significant impact on human ACE2 binding [18]. In the present study, the S gene of 16BO133 (1236 aa) showed a difference of 19 amino acids when compared to SARS CoV (Tor2, 1255 aa) due to 5 aa insertions and 24 aa deletions. Of the 24 aa deletions, 75% (18/24) were located in the RBD. In conclusion, it is thought that 16BO133 may have very low possibility to human infection due to the mutation of two critical residues (N479S and T487V), two

major deletion sites (433–437, 457–469) in the RBD and low amino acid identity (75.2%) of S gene with SARS CoV Tor2.

According to previous reports [4], B15-21 bat CoV was identified from *R. ferrumequinum* and firstly reported in Republic of Korea. The B15-21 was clustered with the *Betacoronavirus* and grouped with SARS-like bat CoV found in China. The receptor-binding domain (RBD) of B15-21 had two major deletion sites, TGNYN and PFSPDGKPCTPPA, compared to human SARS CoV Tor2. The 16BO133 also had two major deletion sites in RBD, TGNYN (433–437) and NVPFSPDGKPCTP (457–469), compared to human SARS CoV Tor2. The amino acid differences between B15-21 (PFSPDGKPCTPPA) and 16BO133 (NVPFSPDG-KPCTP) are evolving evidence of SARS-like bat CoV in Republic of Korea.

The ORF8 region located upstream of the N gene is known to be a "high mutation region" from previous reports [3, 19]. Most human SARS CoVs during epidemic had undergone 29 nucleotides deletion in ORF8 compared to civet SARS CoV, suggesting that this region may be important for interspecies transmission [20]. In the present study, a complete deletion of amino acids was observed in the ORF8 region of 16BO133. Interestingly, insertion of two nucleotides (cytosine and threonine) was observed in front of the stop codon of ORF7b. The insertion of two nucleotides induced an ORF frame shift resulting in addition of four amino acids of ORF7b and an elimination of the start codon of ORF8. Further studies are needed on how these changes will influence SARS-like bat CoV.

According to previous reports, SARS-like bat CoV (RP3) was first discovered in China [19]. The overall sequence identity between RP3 and human SARS CoV Tor2 was

92%. However, the S1 domain of the S protein showed 64% sequence identity due to amino acid deletions. After the discovery of RP3, two novel SARS-like bat CoVs (Rs3367 and LYRa11) have been described, which are more closely related to human SARS CoV Tor2 [7, 20]. Rs3367 and LYRa11 have high amino acid identities of 89.6% to 89.9%, respectively, with human SARS CoV Tor2, particularly in the RBM region without amino acid deletion. The evolution of the CoV can lead to a novel CoV that is highly contagious in humans, which can lead to a serious problem.

In conclusion, the CoV can possibly be transmitted to human populations due to CoV mutations occurring as a result of high mutation rates as the virus evolves. Therefore, continuous monitoring and genomic sequence characterization of the SARS-like bat CoV should be performed to prevent human infections that may result from genetic variation.

Compliance with ethical standards

Conflict of interest The authors declare that there are no conflicts of interest.

Ethical approval This article did not involve studies of human subjects or animals.

References

1. Brian DA, Baric RS (2005) Coronavirus genome structure and replication. Curr Top Microbiol Immunol 287:1–30
2. Gonzalez JM, Gomez-Puertas P, Cavanagh D, Gorbalenya AE, Enjuanes L (2003) A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. Arch Virol 148:2207–2235
3. Xu L, Zhang F, Yang W, Jiang T, Lu G, He B, Li X, Hu T, Chen G, Feng Y, Zhang Y, Fan Q, Feng J, Zhang H, Tu C (2016) Detection and characterization of diverse alpha- and betacoronaviruses from bats in China. Virol Sin 31:69–77
4. Kim HK, Yoon SW, Kim DJ, Koo BS, Noh JY, Kim JH, Choi YG, Na W, Chang KT, Song D, Jeong DG (2016) Detection of severe acute respiratory syndrome-like, Middle East respiratory syndrome-like bat coronaviruses and group H rotavirus in faeces of Korean bats. Transbound Emerg Dis 63:365–372
5. Yang L, Wu Z, Ren X, Yang F, He G, Zhang J, Dong J, Sun L, Zhu Y, Du J, Zhang S, Jin Q (2011) Novel SARS-like betacoronaviruses in bats, China. Emerg Infect Dis 19:989–991
6. Ren W, Li W, Yu M, Hao P, Zhang Y, Zhou P, Zhang S, Zhao G, Zhong Y, Wang S, Wang LF, Shi Z (2006) Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. J Gen Virol 87:3355–3359
7. He B, Zhang Y, Xu L, Yang W, Yang F, Feng Y, Xia L, Zhou J, Zhen W, Feng Y, Guo H, Zhang H, Tu C (2014) Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. J Virol 88:7070–7082
8. Nuttall I, Dye C (2013) Epidemiology. The SARS wake-up call. Science 339:1287–1288
9. Peiris JS, Yuen KY, Osterhaus AD, Stohr K (2003) The severe acute respiratory syndrome. N Engl J Med 349:2431–2441
10. Li F, Li W, Farzan M, Harrison SC (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309:1864–1868
11. Sui J, Li W, Murakami A, Tamin A, Matthews LJ, Wong SK, Moore MJ, Tallarico AS, Olurinde M, Choe H, Anderson LJ, Bellini WJ, Farzan M, Marasco WA (2004) Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. Proc Natl Acad Sci U S A 101:2536–2541
12. Qu XX, Hao P, Song XJ, Jiang SM, Liu YX, Wang PG, Rao X, Song HD, Wang SY, Zuo Y, Zheng AH, Luo M, Wang HL, Deng F, Wang HZ, Hu ZH, Ding MX, Zhao GP, Deng HK (2005) Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. J Biol Chem 280:29588–29595
13. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A 102:14040–14045
14. Lee S, Jo SD, Son K, An I, Jeong J, Wang SJ, Kim Y, Jheong W, Oem JK (2018) Genetic characteristics of coronaviruses from Korean bats in 2016. Microb Ecol 75:174–182
15. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong SK, Huang IC, Xu K, Vasilieva N, Murakami A, He Y, Marasco WA, Guan Y, Choe H, Farzan M (2005) Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J 24:1634–1643
16. Belouzard S, Chu VC, Whittaker GR (2009) Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc Natl Acad Sci U S A 106:5871–5876
17. Spiga O, Bernini A, Ciutti A, Chiellini S, Menciassi N, Finetti F, Causarono V, Anselmi F, Prischi F, Niccolai N (2003) Molecular modelling of S1 and S2 subunits of SARS coronavirus spike glycoprotein. Biochem Biophys Res Commun 310:78–83
18. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang YJ, Luo CM, Tan B, Wang N, Zhu Y, Crameri G, Zhang SY, Wang LF, Daszak P, Shi ZL (2013) Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503:535–538
19. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang LF (2005) Bats are natural reservoirs of SARS-like coronaviruses. Science 310:676–679
20. Lau SK, Feng Y, Chen H, Luk HK, Yang WH, Li KS, Zhang YZ, Huang Y, Song ZZ, Chow WN, Fan RY, Ahmed SS, Yeung HC, Lam CS, Cai JP, Wong SS, Chan JF, Yuen KY, Zhang HL, Woo PC (2015) Severe Acute Respiratory Syndrome (SARS) coronavirus ORF8 protein is acquired from SARS-related coronavirus from greater horseshoe bats through recombination. J Virol 89:10532–10547