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.
Review

Molecular epidemiology, evolution and phylogeny of SARS coronavirus

Hayes K.H. Luk a, Xin Li a, Joshua Fung a, Susanna K.P. Lau a,b,c,d,⁎, Patrick C.Y. Woo a,b,c,d,⁎

a Department of Microbiology, The University of Hong Kong, Hong Kong, China
b State Key Laboratory of Emerging Infectious Diseases, The University of Hong Kong, Hong Kong, China
c Carol Yu Centre for Infection, The University of Hong Kong, Hong Kong, China
d Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University, Hangzhou 310006, China

ABSTRACT

Shortly after its emergence in southern China in 2002/2003, Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) was confirmed to be the cause of SARS. Subsequently, SARS-related CoVs (SARSr-CoVs) were found in palm civets from live animal markets in Guangdong and in various horseshoe bat species, which were believed to be the ultimate reservoir of SARSr-CoV. Till November 2018, 339 SARSr-CoV genomes have been sequenced, including 274 from human, 18 from civets and 47 from bats [mostly from Chinese horseshoe bats (Rhinolophus sinicus), n = 30; and greater horseshoe bats (Rhinolophus ferrumequinum), n = 9]. The human SARS-CoVs and civet SARSr-CoVs were collected in 2003/2004, while bat SARSr-CoVs were continuously isolated in the past 13 years even after the cessation of the SARS epidemic. SARSr-CoVs belong to the subgenus Sarbecovirus (previously lineage B) of genus Betacoronavirus and occupy a unique phylogenetic position. Overall, it is observed that the SARSr-CoV genomes from bats in Yunnan province of China possess the highest nucleotide identity to those from civets. It is evident from both multiple alignment and phylogenetic analyses that some genes of a particular SARSr-CoV from bats may possess higher while other genes possess much lower nucleotide identity to the corresponding genes of SARSr-CoV from human/civets, resulting in the shift of phylogenetic position in different phylogenetic trees. Our current model on the origin of SARS is that the human SARS-CoV that caused the epidemic in 2002/2003 was probably a result of multiple recombination events from a number of SARSr-CoV ancestors in different horseshoe bat species.

1. Introduction

Shortly after its emergence in southern China in 2002/2003, Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) was confirmed to be the cause of SARS (Peiris et al., 2003). There has been a total of 8096 laboratory-confirmed cases of SARS, leading to 774 mortalities in 11 countries (World Health Organization, 2004). Subsequently, SARS-related CoV (SARSr-CoV) was found in palm civets from live animal markets in Guangdong province (Guan et al., 2003). However, the initial hypothesis that civets may act as animal reservoir of SARS-CoV was soon overturned by several observations. Firstly, SARSr-CoV was only detected in civets from the market, but not those in the wild (Kan et al., 2005; Tu et al., 2004). Secondly, the high ratio of nonsynonymous to synonymous mutation rates (Ka/Ks ratios) of the Spike (S), open reading frame (ORF) 3a and non-structural protein (nsp) genes in civet SARS-CoVs collected in both the 2003 and the minor 2004 outbreaks suggested that the virus was undergoing rapid evolutionary gene adaptation in civets (Song et al., 2005). Thirdly, compared with SARS-CoV collected from human during the 2003 epidemic, functional changes have been observed in the S protein of civet SARS-CoV and the SARS-CoV isolated from the 2004 minor outbreak. The latter showed less efficient use of the human angiotensin converting enzyme 2 (ACE2) receptor (Li et al., 2005b) and demonstrated resistance to antibody inhibition (Yang et al., 2005). Finally, while significant levels of antibody to SARS-CoV were detected in 80% of the civets from one animal market in Guangzhou, low seroprevalence rates in civets from various civet farms in China suggested that civets were largely brought to the animal market infection-free (Guan et al., 2003; Tu et al., 2004). The SARSr-CoV was likely contracted during later mixing and trading.

In view of these observations, we carried out a molecular surveillance study in various mammals in Hong Kong to hunt for the ultimate source of the virus (Lau et al., 2005). Among the 127 bats (including 8 bat species), 20 monkeys and 60 rodents surveyed, SARSr-CoV was only detected in 39% of the fecal samples of 59 Chinese horseshoe bats...
(Rhinolophus sinicus) (Lau et al., 2005). Western blot analysis showed that antibodies against the nucleocapsid (N) protein of bat SARSr-CoV was present in 67% of the serum samples of Chinese horseshoe bats, while 8% of the serum of Chinese horseshoe bats were tested positive for human SARS-CoV-neutralizing antibody with titer \( \geq 1:20 \) (Lau et al., 2005). Shortly afterward, another independent group also reported the detection of SARSr-CoV in Chinese horseshoe bats, greater horseshoe bats (Rhinolophus ferrumequinum, RF), and big-eared horseshoe bats (Rhinolophus macrozonus, Rm) in Hubei and Guangxi provinces of China (Li et al., 2005a). In the past few years, SARSr-CoVs have been isolated from a variety of horseshoe bats in Yunnan province of China by several groups (Ge et al., 2013; He et al., 2014; Hu et al., 2017; Lau et al., 2015).

Three hundred and thirty nine SARSr-CoV genomes have been sequenced from 2003 to 2018. These include 274 genomes from human, 18 from civets and 47 from bats. The human SARS-CoV and civet SARSr-CoV were collected in 2003/2004, while bat SARSr-CoVs were continuously detected even after the cessation of the SARS epidemic. In this article, we review our current understanding of the molecular epidemiology, evolution and, phylogeny of SARSr-CoVs based on analysis of these 339 genomes.

2. The SARS-CoV genome

The genome size of the SARS-CoV varies from 29.0 kb to 30.2 kb. Its genome structure follows the characteristic gene order of other known CoVs: the 5’ two thirds of the genome comprises ORF1ab encoding replicase polyproteins, while the 3’ one third consists of genes encoding structural proteins including S, envelope (E), membrane (M), and N proteins (Fig. 1). Both the 5’ and 3’ ends of the SARS-CoV genome contain short untranslated regions. The translational product of ORF1ab is cleaved by proteases encoded by SARSr-CoV itself into 16 nsps, which include major enzymes such as papain-like protease(s) (PLpro), chymotrypsin-like protease (3CLpro), RNA-dependent RNA polymerase (RdRp) and helicase (Hel) (Fig. 1). In contrast to the genome of viruses belonging to lineage A Betacoronavirus (recently renamed as subgenus Embecovirus), haemagglutinin-esterase gene is absent from the genome of SARS-CoV. In addition, SARS-CoV contains 6–7 accessory proteins, encoded by ORF3a, ORF3b, ORF6, ORF7a, ORF7b and ORF8 (or ORF8a and ORF8b as a result of a 29-nucleotide (nt) deletion). This is unique to lineage B Betacoronavirus, a subgenus recently renamed Sarbecovirus and contains all SARSr-CoVs (Fig. 2).

Studies in the past 15 years have partly revealed the biochemical functions of these accessory proteins (Liu et al., 2014). Protein 3a triggers apoptosis and induces the production of proinflammatory cytokines such as RANTES (Regulated on Activation, Normal T Cell Expressed and Secreted; also known as C-C motif chemokine ligand 5, CCL5) and CXCL8 (C-X-C motif chemokine ligand 8). Protein 3b inhibits type I interferon and also induces apoptosis. Protein 6 inhibits interferon signaling and stimulates DNA synthesis. Protein 7a activates NF-kB (nuclear factor kappa B) and MAPK8 (mitogen-activated protein kinase 8) for CXCL8 and RANTES production. The function of protein 7b is not well characterized yet. ORF8 is present in all SARSr-CoV genomes in bats and civets, as well as in SARS-CoVs isolated from human during the early phase of the epidemic. Protein 8 activates the ATF6 (activating transcription factor 6) branch of unfolded protein response. In the genomes of SARS-CoVs isolated from human during the late phase of the epidemic, there was a signature 29-nt deletion in ORF8, splitting it into two separate ORFs 8a and 8b. Protein 8a includes caspase-dependent apoptosis whereas protein 8b modulates cellular DNA synthesis.

3. Unique phylogenetic position of SARS-CoV

Before the SARS epidemic, there were just around 10 CoVs with complete genome sequences available. These CoVs were classified into three groups: Group 1, Group 2 and Group 3. In 2011, the Coronavirus Study Group of the International Committee for Taxonomy of Viruses renamed these three groups into three genera: Alphacoronavirus, Betacoronavirus, and Gammacoronavirus (de Groot et al., 2011). When SARS-CoV was first discovered in 2003, phylogenetic analysis of the SARS-CoV genome showed that it occupied a unique position in Betacoronavirus, which was subsequently placed into the subgenus Sarbecovirus. The traditional betaCoVs (e.g. mouse hepatitis virus, human CoV OC43, bovine CoV) were classified as Embecovirus (Fig. 2). After the SARS epidemic, an unprecedented number of novel CoVs were discovered (Lau et al., 2016; Lau et al., 2007; Lau et al., 2012a; Lau et al., 2012b; Lau et al., 2014; Woo et al., 2005; Woo et al., 2014a; Woo et al., 2014b). This led to the description of lineage C Betacoronavirus, which comprises important members such as Tylonycteris bat coronavirus HKU4, Pipistrellus bat coronavirus HKU5, Hypsugo bat coronavirus HKU25 and Middle East Respiratory Syndrome CoV (Lau et al., 2013; Lau et al., 2018b; Woo et al., 2007; Woo et al., 2006a), and lineage D Betacoronavirus (Lau et al., 2010b) as well as a novel genus Deltacoronavirus (Lau et al., 2015a; Woo et al., 2012; Woo et al., 2017) (Fig. 2). Lineage C and lineage D Betacoronavirus were now renamed as subgenera Merbecovirus and Nobecovirus.

4. Molecular epidemiology and evolution of SARS-CoV

4.1. Circulation of SARSr-CoV in horseshoe bats in 2004 to 2018

Since its first discovery in Chinese horseshoe bats in 2004 (Lau et al., 2005), SARSr-CoVs have been continuously found in various horseshoe bat species in the last 13 years (Drexler et al., 2010; Ge et al., 2013; He et al., 2014; Hu et al., 2017; Lau et al., 2005; Lau et al., 2010a; Li et al., 2005a; Tang et al., 2006; Wu et al., 2016; Yang et al., 2013; Yang et al., 2015; Zeng et al., 2016). This is in contrast to the case for civets and human, where SARS-CoVs were only found in 2003/2004, and never reported afterward. For the 47 bat SARSr-CoV genomes, 30 are from Chinese horseshoe bats, 9 from greater horseshoe bats, 2 from big-eared horseshoe bat, 2 from least horseshoe bat (Rhinolophus pusillus, Rp), and 1 each from intermediate horseshoe bat (Rhinolophus affinis, Ra), Blasiu’s horseshoe bat (Rhinolophus blasii, Rb), Stoliczka’s Asian trident bat (Aselliscus stoliczkanus, As) in the neighboring family Hipposideridae and wrinkled-lipped free-tailed bat (Chaerephon plicatus, Cp) in the genetically more distant family Molossidae. SARSr-CoVs have also been detected in countries other than China, including Thailand, Italy, Luxembourg, Bulgaria, Slovenia, Hungary, Japan, Kenya, etc. However, only partial sequences were available for these isolates. Nevertheless, the immediate progenitor of SARS-CoV has not been pinpointed.

---

Fig. 1. Genome organization of SARS-CoV. ORF1ab with nsp1–16 are colored in blue. Structural proteins including S, E, M and N are in pink. Accessory proteins were numbered and in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Among the 45 SARS-CoV genomes from bats in China, 11 of them were from Hong Kong, 2 were from Guangdong, 2 from Guangxi, 5 from Hubei, 20 from Yunnan and one each from Shaanxi, Shanxi, Jilin, Guizhou, Hebei, respectively. Overall, it is observed that the SARS-CoV genomes from bats in Yunnan possess relatively higher nt identity to those from civets (Fig. 3A). This mismatch between clinical events and apparent gradient of nt identity could either be due to a missing link during evolutionary adaptation among the SARS-CoVs in different provinces or simply as a result of sampling error.

4.3. Recombination and evolution

The high frequency of homologous RNA recombination is one of the major factors contributing to a plastic genome underpinning the evolutionary force in CoVs. This has resulted in different genotypes or even different CoVs adapted to new hosts (Herrewegh et al., 1998; Lau et al., 2011; Terada et al., 2014; Woo et al., 2006b). As for SARSr-CoVs, it is
evident from both multiple alignments and phylogenetic analyses that some genes of a particular SARSr-CoV from bats may possess higher while other genes possess much lower nt identity to the corresponding genes of SARSr-CoV genomes from human/civets, resulting in shifting of phylogenetic position in different phylogenetic trees. This phenomenon is frequently observed in SARSr-CoVs and likely explains the generation of novel SARSr-CoV strains that could jump from bat to civet and subsequently to human.

4.4. S protein of SARSr-BatCoVs

Trimers of S protein form spikes on the surface of CoV particles. It comprises two functionally distinct subunits – S1 and S2 domains which are involved in receptor binding and fusion respectively. Like other class I viral fusion proteins, the S protein undergoes a series of events including receptor recognition, proteolytic cleavage to shed the S1 subunit and conformational changes in S2 that ultimately lead to fusion of the viral and host membranes. It has been well established that human SARS-CoV utilizes ACE2 as a functional receptor. The receptor-binding motifs in the C-terminal domain of S1 are implicated in receptor recognition. Substitutions within the S1 receptor-binding domain (RBD) confers adaptability to new or orthologous entry receptors, thus altering the viral tropism (Hulswit et al., 2016). Undoubtedly, the ability to bind human ACE2 is an indispensable step in establishing cross-species transmission.

As we have previously hypothesized, the five amino acid (a.a.) deletion, twelve a.a. deletion as well as a.a. substitutions at 5 critical residues for binding serve as determining factors for the S protein-ACE2 interaction (Table 1). A critical pre-requisite for reliable ACE2 utilization is the absence of 5 a.a. and 12 a.a. deletions and, preferably, the presence of at least two out of five human-adapted residues. Based on these analyses, SARSr-Rs-BatCoV WIV1 is one of the strains most advantageously conformed genotypically for ACE2 utilization. It has been shown to be able to directly infect well-differentiated primary human airway epithelial cell cultures (Menachery et al., 2016). In addition, neutralization assays using convalescent sera from SARS-CoV infected patients showed robust neutralization against tissue culture infectious dose 50 of WIV1 (Ge et al., 2013).

SARSr-Rs-BatCoV RsSHC014, a strain discovered in Chinese horseshoe bats in 2012, contains two of the five a.a. residues in civet strain civet007 but none of the five human-adapted residues. It retains both the 5 a.a. and 12 a.a. deletion sites. This genotype is shared by SARSr-Rs-BatCoV Rs4231 and Rs4084. Recombinant mouse-adapted SARS-CoV expressing the S protein of RsSHC014 was still able to utilize ACE2 for viral entry, causing cytopathic changes in Vero cells and weight loss in mice model, despite the apparent failure in pseudovirus infectivity assay (Menachery et al., 2015). However, significantly slower viral replication rate was observed, suggesting that deletions in RBDs were more critical in receptor recognition while the presence of ACE2-adapted critical residues modulated entry efficiency. In contrary to the WIV1 strain, neither neutralizing human monoclonal antibody nor existing double-inactivated whole SARS-CoV vaccine provided protective effect against infection caused by SHC014-harboring recombinant virus strain, indicating key difference in a.a. sequences determining antigenicity (Menachery et al., 2015).

A phylogenetic tree was constructed from the RBDs of SARSr-CoVs (Fig. 4). Surprisingly, SARSr-Ra-BatCoV LyRa11, a strain discovered in intermediate horseshoe bat in Baoshan, Yunnan (He et al., 2014), around 375 km from Kunming, sits closest to the civet/human SARSr-CoVs despite its lower overall nt identity. This is likely due to the presence of 7 nt in civet/human SARS-CoV that is found exclusively in SARSr-Ra-BatCoV LyRa11. Such a unique S gene in relation to civet/human SARS-CoV would therefore challenge the previous hypothesis that the origin of RBD was solely from Chinese horseshoe bats in

![Fig. 3.](image-url)
Table 1

| SARSr-BatCoVs that could replicate in cell lines | SARSr-BatCoVs that were reported to have slower replication kinetics (Menachery et al. 2015) | SARSr-BatCoVs that failed to replicate in cell lines |
|-----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| Human SARS-CoV TOR2 | * Critical a.a. residues on S protein determining interaction with ACE2 | * Only RBD sequences are available in Genbank |

Table 1

Summary of the critical elements for ACE2 utilization present in the S protein of various SARSr-BatCoVs.

| SARSr-BatCoVs | NTD genotype | 5 a.a. deletion | 12 a.a. deletion | 442^ | 479^ | 491^ |
|---------------|---------------|-----------------|-----------------|-------|-------|-------|
| Human SARS-CoV TOR2 | Retained | Y Retained | L | N | T | Y |
| Civet SARSr-CoV S23 | Retained | Y Retained | L | K | S | Y |
| Civet SARSr-CoV civet007 | Retained | Y Retained | P | R | S | Y |
| SARSr-Rs-BatCoV WIV1 | 2 | Retained | S | Retained | F | N | N | Y |
| SARSr-Rs-BatCoV WIV16 | 1 | Retained | S | Retained | F | N | N | Y |
| SARSr-Rs-BatCoV Rs4874 | 2 | Retained | S | Retained | F | N | N | Y |
| SARSr-Rs-BatCoV LYRa11 | 2 | Retained | S | Retained | F | N | N | Y |
| SARSr-Rs-BatCoV Rs7327 | 2 | Retained | S | Retained | F | N | N | Y |
| SARSr-Rs-BatCoV RsSHC014 | 2 | Retained | W | Retained | P | R | A | H |
| SARSr-Rs-BatCoV Rs4231 | 2 | Retained | W | Retained | P | R | A | H |
| SARSr-Rs-BatCoV Rs4084 | 3 | Deleted | S | Deleted | S | V | Y |
| SARSr-Rs-BatCoV Rs4081 | 2 | Deleted | S | Deleted | S | V | Y |
| SARSr-Rs-BatCoV Rs4075* | 2 | Deleted | S | Deleted | S | P | Y |
| SARSr-Rs-BatCoV Rs4085* | 2 | Deleted | S | Deleted | S | P | Y |
| SARSr-Rf-BatCoV Rf4092 | 3 | Deleted | S | Deleted | S | V | Y |
| SARSr-Rf-BatCoV YNLF 31C | 3 | Deleted | S | Deleted | S | V | Y |

Fig. 3. (continued)
Kunming. Nevertheless, the S proteins of the majority of bat SARSr-CoVs found in Chinese horseshoe bats in various parts of China including Yunnan resemble the genotype of SARSr-Rs-BatCoV HKU3 but not the human ACE2-utilizing genotype of SARSr-Rs-BatCoV WIV1. This raises the question on how SARSr-CoVs manage to have two distinct genotypes of S infecting the same host (Rs). There is a possibility that two cellular receptors for SARSr-CoVs are present in Chinese horseshoe bats. Interestingly, two novel SARSr-Rp-BatCoV strains ZXC21 and ZC45 isolated from least horseshoe bat in Zhoushan, Zhejiang province in eastern China were shown to be able to infect suckling rats, causing inflammation in the brain tissue and histological changes in the lung and intestine despite failed viral isolation in VeroE6 cells. Phylogenetically, these two strains represented a separate clade, lying between the susceptible clade of SARSr-Ra-BatCoV LYRa11 and the non-susceptible clade of SARSr-Rf-BatCoV RF4092, which should have imposed inter-species barrier in terms of receptor specificity based on previous discussion (Hu et al., 2018).

4.5. SARSr-CoV in Chinese horseshoe bats and greater horseshoe bats

Among all the horseshoe bats, SARSr-CoVs are most commonly found in Chinese horseshoe bats in various parts of China including Yunnan. Analysis of their genomes revealed several intriguing phenomena. Firstly, ORF1a and ORF1b in SARSr-Rf-BatCoV YNLF_31C isolated from greater horseshoe bats in Lufeng, Yunnan by our group has the highest nt identity to civet SARSr-CoV, especially at the regions nsp5, nsp10, nsp12 and nsp13 (Fig. 3A). This raised the possibility that Lufeng, Yunnan could be a potential gene pool for the progenitor of ORF1 of SARS-CoV. Secondly, as previously illustrated, for most regions along the genome of SARS-CoV except ORF8, SARSr-Rs-BatCoVs from Chinese horseshoe bats are predominantly closer to civet SARSr-CoV. However, the ORF8 of SARSr-Rf-BatCoV has higher a.a. identity to civet SARSr-CoV ORF8. In addition, V90 and I113 are the two Rf-specific residues present in the ORF8 of SARSr-Rf-BatCoV RF4092, which could evolutionarily bridge the ORF8 of SARSr-Rf-BatCoV to SARSr-Rs-BatCoV that had very high identity to civet/human SARS-CoV (Fig. 5). SARS-Rf-BatCoV was thus hypothesized to be the origin of human SARS-CoV ORF8 gene through recombination events.

In fact, at least three genotypes of ORF8 have been found to circulate in nature (Wu et al., 2016). Type I is genetically closest to ORF8 of civet SARSr-CoV (96.2–98.1% nt identity). There are eleven such strains isolated from bats so far, with eight of them originating from Chinese horseshoe bats and three from greater horseshoe bats. Type II ORF8 has lower (70.8–82.8%) nt identity to ORF8 of civet SARS-CoV. It is prevalent only in SARSr-Rf-BatCoV and demonstrates genetic stability regardless of geographical distribution with similar ORF8s detected in Korea, Shanxi, Hubei, and Hebei. SARSr-Rs-BatCoV has not been found to process type II ORF8. Yet another strain, the SARS-Rp-BatCoV F46 isolated from least horseshoe bats, processes an ORF8 with a phylogenetic position right between type I and type II ORF8s with 93.2% nt identity to civet SARSr-CoV, further complicating the analysis (Fig. 6). The rest of the ORF8 genes from sequenced SARSr-BatCoV strains belong to type III. These type III ORF8 genes possess 53.9–57.7%
nt identity to the ORF8 of civet SARSr-CoV and have the greatest genetic distance observed over the whole genome of SARS-CoV (Fig. 3B). They are detected across a wide range of bat species including R. sinicus, R. macrotris, R. pusillus, R. affinis, A. stolitzkana and C. plicata, except R. ferrumequinum in which type III ORF8 has not been found. In short, the ORF8 of SARSr-Rf-BatCoV from greater horseshoe bats is phylogenetically closer to ORF8 of human/civet SARS-CoV.

As mentioned above, ORF8 in the late-phase human epidemic was split into 8a and 8b due to a unique 29nt deletion. A similar 5nt deletion was observed in the type I ORF8 isolated from the strain SARSr-Rs-BatCoV Rs4084. However, since the majority of SARSr-Rs-BatCoVs’ ORF8 genes belong to the type III ORF8 genotype, it raises the speculation that this minority of type I ORF8 plays a non-essential role in the Chinese horseshoe bats host environment.

In a recent study, recombinant SARS-CoV carrying three forms of ORF8 were constructed, including complete ORF8 as seen in the early phase of the 2003 human epidemic and civets, ORF8 that contained the characteristic 29 nt deletion, and a variant strain with ORF8 replaced by a 5 nt spacer sequence. Replication assay in VeroFM cells demonstrated a phenotypic hierarchy that the infectious clone with full type I ORF8 exhibited significantly more efficient growth than other variants. The same hierarchy of viral replication was observed in human airway epithelial cell cultures and several “non-host” cell lines transduced to express human ACE2, including R. alcyone lung epithelial cell line, cotton rat airway epithelial cell line, and goat and sheep lung cell lines. This study suggests that the 29 nt deletion actually conferred a loss of viral fitness during the initial phase of human-to-human transmission, contradicting the previous belief that the truncated products ORF8a and ORF8b favored adaptation to human (Muth et al., 2018).

4.6. Genomic overview of SARSr-BatCoVs

Complete genome analysis has been carried out for all sequenced SARS-CoV with respect to civet SARS-CoV. The importance of civet SARS-CoV is underlined by its close genetic relationship to human SARS-CoV along the entire genome and especially at the RBD region where nt identity approaches 98.8–99.8%. Moreover, the genetic diversity within the group of civet SARS-CoV strains is minimal, further supporting that civets are only recently infected. Adaptation of SARSr-BatCoV to civet as intermediate amplification host is likely needed, which is further corroborated by the phylogenetic position of civet ACE2 in-between bat and human ACE2. Recently, direct bat-to-human transmission of SARS-CoV was proposed by some groups in view of the fact that some SARSr-BatCoVs could utilize human ACE2 receptors in vitro and that seropositive sera against SARSr-BatCoVs were detected in Yunnan residents (Menachery et al., 2015; Wang et al., 2018), despite the fact that human/civet SARS-CoV has never been reported in Yunnan. Our findings provide substantial evidence against this hypothesis from a phyloepidemiologic point of view and demonstrated the significance of an intermediate host during the transmission cascade (Fig. 3).

Referred to Fig. 3A, SARSr-Rs-BatCoV Rs4084 was one of the two strains that possess the highest genome identity to civet/human SARSr-CoV. It is so far the only SARSr-BatCoV that possessed at least 70% nt identity in all parts of the genome, including the four hypervariable regions. The region sharing lowest similarity to civet/human SARSr-CoV was the N-terminal domain (NDT) of the S protein which was likely non-critical. As mentioned above, SARSr-Rs-BatCoV Rs4084 was predicted to be able to utilize human ACE2 as functional receptor since its RBD region is genetically similar to RsSHC014. SARSr-Rs-BatCoV Rs4084 has not been studied either as whole virus or in the form of recombinant virus in terms of accessory protein function, tissue infectivity assay and replication kinetics; leaving a potential field for future investigations. Animal experiment using civet model would offer great promise in the studying of such virus, along with other SARSr-BatCoVs that could utilize human ACE2.

5. Concluding remarks

The > 300 genome sequences of SARS-CoV in human, civets and bats accumulated in the last 15 years allowed us to understand the cause of such a highly fatal epidemic that affected us in a global scale in 2003. After 15 years of work, we are now much closer to fully understand the origin of SARS-CoV and its evolution, as genomes with fragments that contain gene sequences with higher and higher nt identities to the SARS-CoV found in human were observed. Our current model on the origin of SARS is that the human SARS-CoV that caused the epidemic in 2003 was probably a result of multiple recombination events from a number of SARS-CoV ancestors. Yunnan is the province in China with the largest diversity of horseshoe bat species. It is also the area shown to have a high variety of SARS-CoV ancestors. Further in-depth molecular epidemiology studies in this and other provinces in its proximity, including Guangxi and Guangdong, will hopefully give us an even clearer picture on the origin and early evolution of SARS-CoV.

Acknowledgements

This work is partly supported by the Theme-based Research Scheme (Project No. T11/707/15), University Grant Committee; Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Diseases and Research Capability on Antimicrobial Resistance for Department of Health, HK SAR Government; and University Development Fund, The University of Hong Kong.

References

Campanella, J.J., Bittincka, L., Smallley, J., 2003. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4, 29.

de Groot, R.J., Baker, S.C., Baric, R., Enjuanes, L., Gorbalenya, A., Holmes, K.V., Perlman, S., Poony, L., Rottier, P.J., Taillon, P.J., Tovo, P., Zinkah, J., 2011. Coronavirus. In: Virus Taxonomy, Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses. International Union of Microbiological Societies, Virology Division, pp. 806–828.

Drewer, J.F., Glinz-Bauch, F., Glinz, J., Corman, V.M., Muth, D., Goettische, M., Seebens, A., Niedrig, M., Pfefferle, S., Vorderwurmd, S., Zehlyakov, Z., Hermanns, U., Vallo, P., Lukashe, A., Muller, M.A., Deng, H., Herrler, G., Drosten, C., 2010. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. J. Virol. 84, 11336–11349.

Gu, X.Y., Li, J.J., Yang, X.L., Chumara, A.A., Zhu, G., Epstein, J.H., Maet, J.K., Hu, B., Zhang, W., Peng, C., Zhang, Y.J., Luo, C.M., Tan, B., Wang, N., Zhu, Y., Cramer, G., Zhang, S.Y., Wang, L.F., Daszak, P., Shi, Z.L., 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503, 535–538.

Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhang, Z.X., Cheng, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., Butt, K.M., Wong, K.L., Chan, K.W., Lim, W., Shortridge, K.F., Yuen, K.Y., Peiris, J.S., Poon, L.L., 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302, 276–278.

Herrewegh, A.A., Smeenk, I., Horzinek, M.C., Rottier, P.J., de Groot, R.J., 1998. Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. J. Virol. 72, 4508–4514.

Hu, B., Zeng, L.P., Yang, X.L., Ge, X.Y., Zhang, W., Li, B., Xie, J.Z., Shen, Y.R., Zhang, Y.Z., Wang, N., Luo, D.S., Zheng, X.S., Wang, M.N., Daszak, P., Wang, L.F., Cui, J., Shi,
Liu, D.X., Fung, T.S., Chong, K.K., Shukla, A., Hilgenfeld, R., 2014. Accessory proteins of a novel SARS-like coronavirus in bats. Emer. Microbes Infect. 3, 154.

Hulstaert, R.J., de Haan, C.A., Bosch, B.J., 2016. Coronavirus spike protein and tropism determinants. J. Virol. Methods 240, 29–37.

Kan, B., Wang, M., Jing, H., Xu, H., Jiang, X., Yan, M., Liang, W., Zheng, H., Wan, K., Liu, Q., Cai, B., Xu, Y., Zhang, E., Wang, H., Ye, J., Li, G., Li, M., Cui, Z., Qi, X., Chen, D., Wu, Y., Yang, J., Wu, X., Jia, X., Xu, J., Cai, S., 2005. Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm cirivets at an animal market in Guangdong, China. J Virol. 79, 11923–11932.

Lau, S.K., Wong, P.C.Y., Lau, S.K.P., Poon, R.W.S., Wong, B.H.L., Tsoi, H.-w., Yip, B.C.K., Lau, J.H., Bai, R., Teng, J.L., Tsang, A.K., Lau, C.C., 2012b. Recent transmission of a novel Alphacoronavirus bat HKU4 from horseshoe bats to man. J Virol. 86, 5481–5496.

Lau, S.K., Li, K.S.M., Huang, Y., Shek, C.T., Leung, M., Chai, G.K.Y., Xu, H., Lam, C.S.F., Guo, R., Chan, K.H., Zhong, B.J., Woo, P.C.Y., Yuen, K.Y., 2010a. Ectodomain and complete genome comparison of different strains of severe acute respiratory syndrome coronavirus. J Virol. 84, 428–439.

Lau, S.K.P., Li, K.S.M., Huang, Y., Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.K., Huang, I.C., Xu, K., Lau, S.K., Woo, P.C., Yip, C.C., Fan, R.Y.Y., Luk, H.K.H., Cai, J.-P., Lau, S.K., Woo, P.C., Yip, C.C., Fan, R.Y., Huang, Y., Wang, M., Guo, R., Li, K.S.M., Lau, S.K., Wu, P.C., Cai, J.-P., Chan, J.H.K., Cheung, B.T.Y., Zhou, H.-Q., Xiang, H., Zheng, H., Jhang, J.H., Chern, S.W., Cheng, D., Pan, C.M., Xuan, H., Chen, S.J., Luo, H.M., Zhou, D.H., Liu, Y.F., He, J., Qin, P.Z., Li, H., Ren, Y.Q., Tang, W.J., Yu, Y.D., Anderson, L., Wang, M., Xu, R.H., Wu, X.W., Zheng, H., Chen, J.D., Liang, G., Yao, L., Miao, F., Liang, Y.L., Li, H., Chen, F., Li, H., He, L.J., Lin, J.Y., Tong, S., Kong, X., Lu, D., Hao, P., Tang, H., Bernimi, A., Yu, X.J., Spiga, O., Guo, Z.M., Fan, H.Y., He, W.Z., Manuguerra, J.C., Fontanet, A., Danchin, A., Nicolai, L., Ni, Y.X., Wu, C.L., Zhao, G.P., 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc. Natl. Acad. Sci. U.S.A. 102, 4340–4345.

Yuen, K.Y., 2013. Genetic characterization of Betacoronavirus lineage C viruses in bats reveals a reservoir for acute, self-limiting infection that allows recombination events. J Virol. 87, 20888–20891.

Yang, X.L., Hu, D., Zuo, Y., Wang, B.H.L., Wang, M., Huang, Y., Xu, H., Guo, R., Li, K.S.M., Guo, K., Chan, K.H., Zhong, B.J., Woo, P.C.Y., Yuen, K.Y., 2010. Coexistence of different genotypes in the same bat and serological characterization of Rousettus bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. J Virol. 84, 11325–11334.

Lau, S.K., Lee, P., Tsang, A.K.L., Yip, C.C.Y., Shek, C.T., Leung, M., Cui, Z., Chen, D., Wu, X., Chim, W.C., Fan, Y., Lai, M., Chi, K.H., Lu, L., Cai, J.P., Chan, J.H.K., Cheung, B.T.Y., Zhou, H.-Q., Xiang, H., Zheng, H., Jhang, J.H., Chern, S.W., Cheng, D., Pan, C.M., Xuan, H., Chen, S.J., Luo, H.M., Zhou, D.H., Liu, Y.F., He, J., Qin, P.Z., Li, H., Ren, Y.Q., Tang, W.J., Yu, Y.D., Anderson, L., Wang, M., Xu, R.H., Wu, X.W., Zheng, H., Chen, J.D., Liang, G., Yao, L., Miao, F., Liang, Y.L., Li, H., Chen, F., Li, H., He, L.J., Lin, J.Y., Tong, S., Kong, X., Lu, D., Hao, P., Tang, H., Bernimi, A., Yu, X.J., Spiga, O., Guo, Z.M., Fan, H.Y., He, W.Z., Manuguerra, J.C., Fontanet, A., Danchin, A., Nicolai, L., Ni, Y.X., Wu, C.L., Zhao, G.P., 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc. Natl. Acad. Sci. U.S.A. 102, 4340–4345.