



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.



ELSEVIER

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

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

ARTICLE INFO

Keywords:

SARS coronavirus
Molecular epidemiology
Evolution
Phylogeny

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 SARSr-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 SARSr-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

* Corresponding authors at: State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, Block T, Queen Mary Hospital, Hong Kong, China.

E-mail addresses: skplau@hku.hk (S.K.P. Lau), pcywoo@hku.hk (P.C.Y. Woo).

<https://doi.org/10.1016/j.meegid.2019.03.001>

Received 11 December 2018; Received in revised form 27 February 2019; Accepted 1 March 2019

Available online 04 March 2019

1567-1348/© 2019 Published by Elsevier B.V.

(*Rhinolophus sinicus*, Rs) (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 macrotis*, 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 SARS-CoV itself into 16 nsp, which include major enzymes such as papain-like protease(s) (PL^{pro}), chymotrypsin-like protease (3CL^{pro}), 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 subgenera *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- κ B (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., 2018a; 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 SARSr-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), Blasius'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 plicata*, 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.

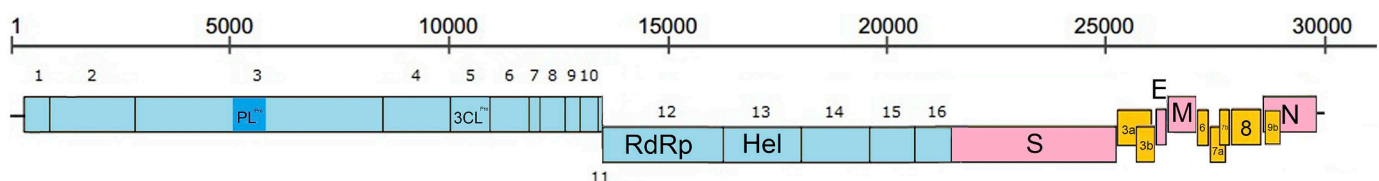


Fig. 1. Genome organization of SARS-CoV. ORF1ab with nsp1–16 are colored in blue. Structural proteins including S, E, M and N are in pink. Accessory proteins were numbered and in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

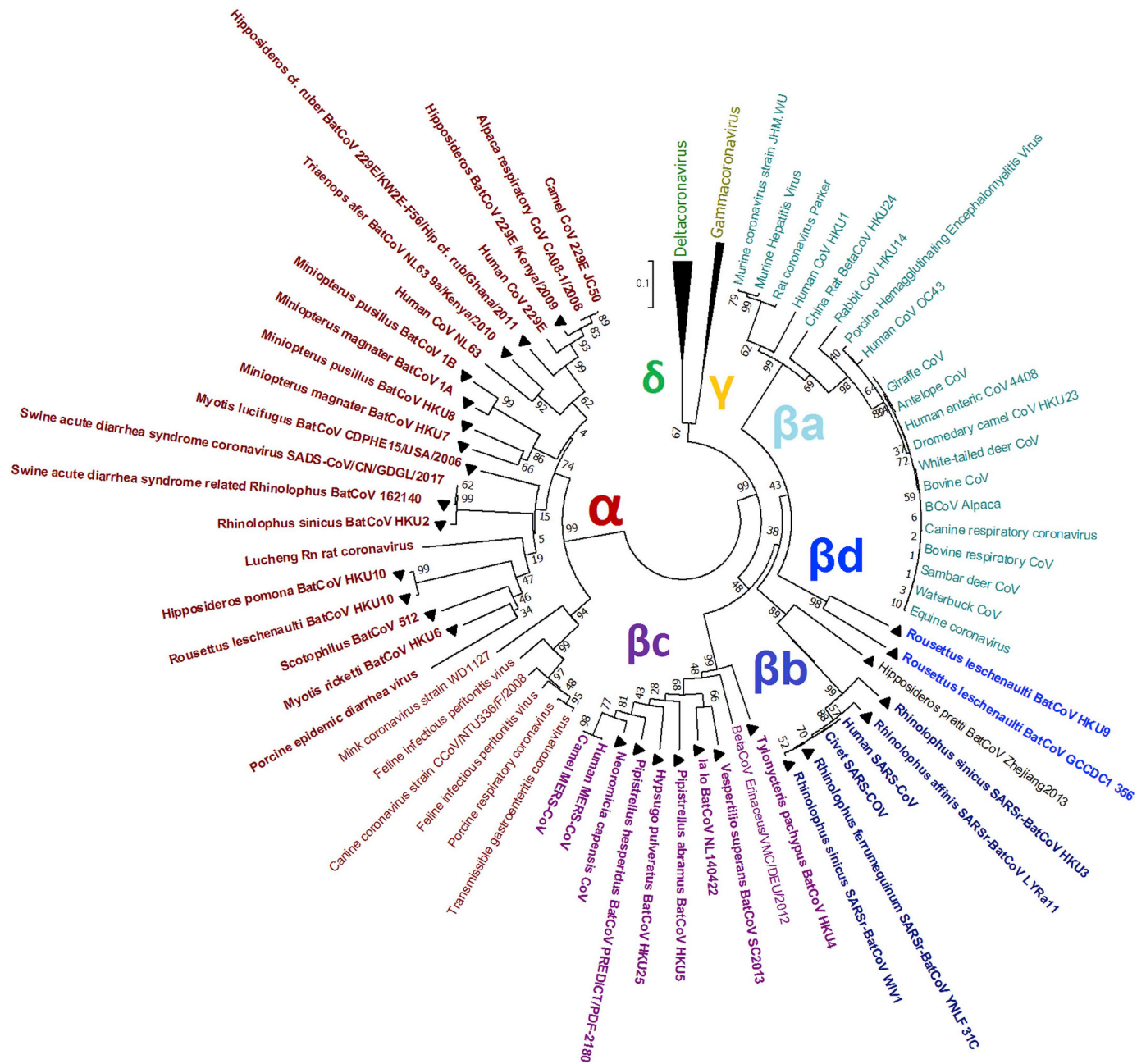


Fig. 2. The diversity of CoVs as demonstrated with a phylogenetic tree targeting the 303 bp partial RdRp sequence (position 15,293–15,596 with respect to Human SARS-CoV TOR2). The neighbor-joining phylogenetic tree was constructed with maximum composite likelihood method by MEGA 7.0. The test of phylogeny was statistically supported by the bootstraps value calculated from 1000 trees. Bat CoVs were labeled with black triangles. The branches of *Gammacoronavirus* and *Deltacoronavirus* were compressed.

4.2. Geographical gradient of SARSr-CoV

Among the 45 SARSr-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 SARSr-CoV genomes from bats in Yunnan possess relatively higher, whereas those from Guangdong, Jilin, Shanxi and Hebei possess relatively lower nt identity to those from civets (Fig. 3A). This is an interesting phenomenon since the SARS epidemic that emerged in late 2002 was first noticed in Guangdong province as an outbreak of acute community-acquired atypical pneumonia syndrome. Severe cases were later retrospectively traced back to five cities around Guangzhou, with the index case reported in Foshan, a city 24 km away from Guangzhou. The

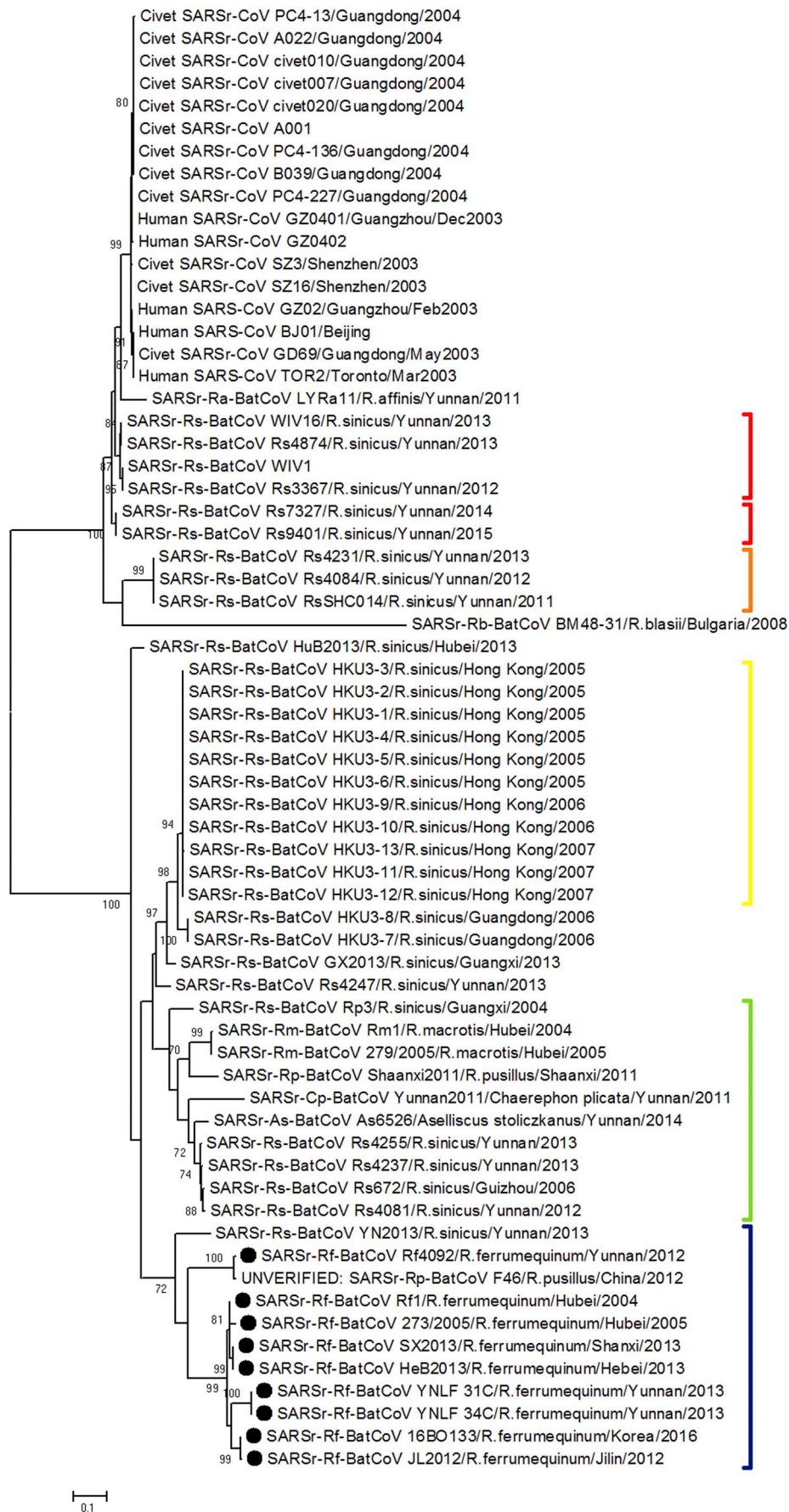
second case was a chef from Heyuan who worked in a restaurant in Shenzhen who contacted with wild game-food animals regularly. This mismatch between clinical events and apparent gradient of nt identity could either be due to a missing link during evolutionary adaptation among the SARSr-CoVs in different 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

SARSr-CoV strains	nsp1	nsp2	nsp3	nsp4	nsp5	nsp6	nsp7	nsp8	nsp9	nsp10	nsp11	nsp12	nsp13	nsp14	nsp15	nsp16	NTD	RBD	S2	ORF3	ORF8	E	M	N	Genome		
Civet SARSr-CoV SZ16/Shenzhen2003	99.8	99.8	100	100	100	100	100	100	100	100	100	100	99.8	99.8	99.8	100	100	99.8	99.8	99.8	99.8	99.5	100	99.8	100	99.9	
Human SARSr-CoV GZ02/GuangzhouFeb2003	99.8	99.8	99.9	99.9	99.9	99.9	100	100	100	100	100	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.7	99.1	99.5	100	99.8	100	99.8
Human SARSr-CoV BJ01/Beijing	99.6	99.6	99.9	99.9	99.7	99.9	100	100	100	100	100	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
Human SARSr-CoV GZ02/Mar2003	99.6	99.6	99.9	99.9	99.7	99.9	100	100	100	100	100	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
Civet SARSr-CoV civet020/Guangdong2004	99.5	100	99.7	99.6	99.8	99.3	99.9	100	99.8	100	99.8	99.8	99.3	99.3	99.3	99.3	99.8	99.4	99.1	99.5	99.4	99.5	99.5	100	99.8	99.6	99.5
Civet SARSr-CoV PC4-136/Guangdong2004	99.5	100	99.8	99.6	99.8	99.9	100	99.7	100	99.7	100	99.7	99.7	99.8	100	99.8	99.4	99.1	99.5	99.4	99.5	99.4	99.5	100	99.8	99.7	99.5
Civet SARSr-CoV civet135/Guangdong2004	99.5	100	99.8	99.6	99.8	99.9	100	99.8	100	99.7	100	99.7	99.7	99.8	100	99.8	99.4	99.1	99.5	99.4	99.5	99.4	99.5	100	99.8	99.7	99.5
Human SARSr-CoV GZ2404/GuangzhouDec2003	99.1	100	99.7	99.7	99.9	100	99.9	100	99.8	100	99.7	99.8	99.8	99.8	100	99.8	99.4	99.2	99.1	99.5	99.2	99.5	100	99.8	99.7	99.1	99.1
Civet SARSr-CoV civet007/Guangdong2004	99.1	100	99.7	99.6	99.9	100	99.9	99.6	99.8	100	99.7	99.8	99.8	99.8	100	99.8	99.4	99.2	99.1	99.5	99.2	99.5	100	99.8	99.7	99.1	99.1
Civet SARSr-CoV B039/Guangdong2004	99.1	100	99.7	99.6	99.9	100	99.9	99.6	99.8	100	99.7	99.8	99.8	99.8	100	99.8	99.4	99.2	99.1	99.5	99.2	99.5	100	99.8	99.7	99.1	99.1
Civet SARSr-CoV civet110/Guangdong2004	99.1	100	99.7	99.6	99.9	100	99.9	99.6	99.8	100	99.7	99.8	99.8	99.8	100	99.8	99.4	99.2	99.1	99.5	99.2	99.5	100	99.8	99.7	99.1	99.1
Civet SARSr-CoV A022/Guangdong2004	98.9	100	99.8	99.7	99.9	100	99.9	100	99.8	100	99.8	99.8	99.8	99.8	100	99.7	99.4	99.0	99.4	99.4	99.4	99.5	100	99.8	99.5	98.9	98.9
SARSr-Rs-BatCoV Rs4084/R.sinicus/Yunnan2012	95.7	97.4	96.8	95.9	97.3	98.1	98.9	98.8	98.8	98.8	98.6	98.3	96.6	95.7	93.7	93.5	72.2	81.0	97.2	98.0	96.2	98.7	97.3	98.7	95.7	95.7	
SARSr-Rs-BatCoV Rs4231/R.sinicus/Yunnan2013	95.7	97.4	96.8	95.9	97.3	98.1	98.9	98.8	98.8	98.8	98.6	98.2	96.6	95.7	93.7	93.5	71.9	81.0	97.2	98.0	96.2	98.7	97.3	98.7	95.7	95.7	
SARSr-Rs-BatCoV Rs4018/R.sinicus/Yunnan2015	95.6	97.4	96.8	95.9	97.3	98.6	99	98.8	98.7	98.8	98.6	98.2	96.6	95.7	93.7	93.5	71.9	81.0	97.2	98.0	96.2	98.7	97.3	98.7	95.6	95.6	
SARSr-Rs-BatCoV Rs3367/R.sinicus/Yunnan2012	95.4	97.8	96.4	95.8	97.3	98.7	98.9	98.8	98.8	98.8	98.6	98.3	96.6	95.6	93.6	93.6	72	83.9	97.2	98.0	96.4	99.1	97.3	98.8	95.4	95.4	
SARSr-Rs-BatCoV HSH014/R.sinicus/Yunnan2011	95.2	97.8	96.5	95.6	97.5	98.6	99.1	98.8	98.8	98.8	98.6	98.4	96.9	95.7	93.8	93.4	71.9	81.0	97.4	98.0	96.4	99.1	97.3	98.8	95.4	95.4	
SARSr-Rs-BatCoV WW16/R.sinicus/Yunnan2013	94.3	97.8	96.8	95.5	97.3	98.1	98.7	98.4	98	97.6	97.6	98.1	95.8	96	93.5	93.6	62.4	83.9	97.3	98.0	96.4	99.1	97.3	98.8	94.3	94.3	
SARSr-Rs-BatCoV R4874/R.sinicus/Yunnan2013	94.2	97.6	96.8	95.5	97.3	98.3	98.6	98.4	98	97.6	97.6	98.1	95.5	96.1	93.5	93.6	62.4	83.9	97.3	98.0	96.4	99.1	97.1	98.5	94.2	94.2	
SARSr-Rs-BatCoV Rs7327/R.sinicus/Yunnan2014	93.9	97.6	97	95.8	97.3	98.6	99	98.8	98.7	98.8	98.6	98.1	95.9	95.9	93.4	93	71.9	81.0	97.3	98.0	96.4	99.1	97.1	98.5	93.9	93.9	
SARSr-Rf-BatCoV R4092/R.ferrousquinn/Yunnan2012	93.8	97.6	96.1	95.8	97.3	98.3	99.2	98	98.1	98.4	97.6	98	97.3	95.8	93.1	92.5	60.5	84.4	97.6	98.0	96.4	99.1	97.1	98.5	93.8	93.8	
SARSr-Rs-BatCoV WIV1	93.8	97.6	96.1	95.8	97.3	98.3	99.2	98	98.1	98.4	97.6	98	97.3	95.8	93.1	92.5	60.5	84.4	97.6	98.0	96.4	99.1	97.1	98.5	93.8	93.8	
SARSr-Rs-BatCoV H4237/R.sinicus/Yunnan2013	93.6	98.8	97.1	95.9	97.6	98.5	99	98.8	98.7	98.8	98.6	98.1	97	95.5	93.5	93.1	61.2	86.6	98.0	96.4	99.1	97.1	98.5	93.8	93.8	93.6	93.6
SARSr-Rs-BatCoV A6526/Aselliscus stoliczkanus/Yunnan2014	93.6	97.6	96.7	95.8	97.3	98.9	98.6	98.4	98.8	98.8	97.9	97.1	98.3	96.7	96.3	93.7	60.7	86.9	98.0	96.4	99.1	97.1	98.5	93.8	93.8	93.6	93.6
SARSr-Rs-BatCoV Rs247/R.sinicus/Yunnan2011	93.5	100	96.8	95.6	99.9	100	99.9	99.6	99.6	100	99.4	99.7	99.7	99.7	99.7	99.7	60.7	86.9	98.0	96.4	99.1	97.1	98.5	93.8	93.8	93.6	93.6
SARSr-Rs-BatCoV R4255/R.sinicus/Yunnan2011	93.4	97.6	96.8	95.5	97.3	98.7	99	98.8	98.8	98.8	98.6	98.2	95.8	96.1	93.6	93.9	62	87	98.5	90.4	94.9	99.7	97.3	98.8	93.4	93.4	
SARSr-Rs-BatCoV R4818/R.sinicus/Yunnan2012	93.4	98	96.8	95.5	97.4	98.3	98.7	98.4	98.1	97.6	97.8	98.2	95.8	95.9	93.6	94.8	62	86.7	98.5	90.5	94.4	99.1	97.1	98.2	93.4	93.4	
SARSr-Rf-BatCoV VNF1_31C/R.ferrousquinn/Yunnan2013	93.2	93.2	93.1	95.7	98.5	96.2	98.4	98.4	98.8	100	97.1	98.6	97.8	96	93.6	93.8	61.7	83.5	98.0	96.4	99.1	97.1	98.5	93.9	93.9	93.2	93.2
SARSr-Rs-BatCoV H4231/R.sinicus/Yunnan2013	93.2	91.1	93	95.7	98.5	96.2	98.4	98.4	98.8	100	97.1	98.6	97.8	96	93.6	93.8	61.7	83.5	98.0	96.4	99.1	97.1	98.5	93.9	93.9	93.2	93.2
SARSr-Rs-BatCoV YN2013/R.sinicus/Yunnan2013	92.3	96.3	96	95.5	97.3	98.6	99.1	98.8	99.1	98.5	98.3	98.2	96.9	95.8	93.8	93.9	64.4	85.1	98.0	96.4	99.1	97.1	98.5	93.9	93.9	92.3	92.3
SARSr-Rs-BatCoV Rp3/R.sinicus/Guangxi2004	92.2	93.5	95.2	93.7	93.8	96.2	98.2	98.4	98.7	98.8	97.4	98	96.5	95.7	94.4	93.9	60.8	86	97.5	98.0	96.4	99.1	97.1	98.5	93.2	93.2	
UNIVERSITY: SARSr-Rp-BatCoV F46/R.pusillus/Yunnan2012	91.9	92.2	87.1	91.3	95.4	94	97.4	95.2	99	97.9	97.6	97.9	98.1	95.6	94.2	93.6	60.5	85.8	98.0	96.4	99.1	97.1	98.5	93.9	93.9	91.9	91.9
SARSr-Rs-BatCoV R072/R.sinicus/Guizhou2006	91.1	97.6	96.2	95.8	96.3	97.9	98.0	97.8	97.8	96.4	97.8	95.8	96.1	93.3	93.8	92	64.4	86.8	98.0	96.4	99.1	97.1	98.5	93.9	93.9	91.1	91.1
SARSr-Rs-BatCoV LYR11/R.rhinis/Yunnan2011	90.8	92.4	87	91.2	90.5	92.3	92.9	90.4	93.1	93.5	94	91.7	91.3	90.5	70.5	68.8	87	90	57.2	88.3	94.6	97	90.8	90.8	90.8	90.8	90.8
SARSr-Rs-BatCoV GX2013/R.sinicus/Guangxi2013	90.5	93.5	88.2	93.2	93.5	97.5	97.7	96.8	98.5	97.1	97.8	97.6	96.5	95.4	94	93.1	80	67.1	86.4	84.1	89.9	98.7	95.5	97.2	90.5	90.5	
SARSr-Cp-BatCoV Yunnan2011/Chaerophora pilicata/Yunnan2011	89.8	92	87.1	95	95.5	92.3	91.2	91.2	93.9	93.2	94.9	92.4	92.1	91.1	89.6	91.6	68.3	86.3	90	57.7	80.7	94.6	91.1	90.8	90.8	89.8	89.8
SARSr-Rs-BatCoV H4201/R.sinicus/Hubei2015	88.6	94.4	89.6	87.1	89.9	91.3	92.5	91.6	93.9	94.1	91.6	92.3	90.5	91	92.1	91.8	60.3	88.6	87.5	83.6	85.4	97.8	95.2	96.4	88.6	88.6	
SARSr-Rm-BatCoV Z792005/R.microtis/Hubei2005	88	92.2	86.1	85.6	90.3	91.1	91.8	92	94.4	94.1	91.8	93.5	90.3	90.3	91.1	91.2	61.8	86.8	87.1	84.1	55.8	97	94.6	95.5	88	88	
SARSr-Rp-BatCoV Shaanxi2011/R.pusillus/Shaanxi2011	87.8	91.5	85.9	85.4	89.9	90.8	91.7	92	94.4	92.9	92.1	93.5	90.2	90	90.5	91.1	61.3	87.6	87.1	83.6	57	96.1	94.6	96	87.8	87.8	
SARSr-Rm-BatCoV Rm1/R.microtis/Hubei2005	87.8	91.5	85.9	85.4	89.9	90.8	91.7	92	94.4	92.9	92.1	93.5	90.2	90	90.5	91.1	61.3	87.6	87.1	83.6	57	96.1	94.6	96	87.8	87.8	
SARSr-Rf-BatCoV R11/R.ferrousquinn/Hubei2004	87.7	91.7	85.3	86.6	89.9	90.7	91.1	92.1	92.6	94.1	91.4	92.5	90.1	89.6	89.1	90.2	64.4	83.9	88.7	83.1	96.5	95.3					

SARSr-CoV strains	nsp1	nsp2	nsp3	nsp4	nsp5	nsp6	nsp7	nsp8	nsp9	nsp10	nsp12	nsp13	nsp14	nsp15	nsp16	NTD	RBD	S2	ORF3	ORF8	E	M	N	Genome	
Civet SARSr-CoV SZ16/Shenzhen/2003	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.001	0.000	0.000	0.000	0.002	0.001	0.001	0.003	0.000	0.003	0.000	0.000	0.001
Human SARSr-CoV G202/GuangzhouFeb2003	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.001	0.001	0.001	0.001	0.004	0.008	0.003	0.010	0.000	0.000	0.005	0.001	0.002
Human SARSr-CoV TOR2/TorontoMar2003	0.000	0.001	0.001	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.000	0.002	0.004	0.011	0.004	0.005	0.000	0.000	0.000	0.000	0.002
Human SARSr-CoV BJ01/Beijing	0.000	0.001	0.001	0.003	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.001	0.000	0.002	0.004	0.011	0.004	0.005	0.000	0.000	0.000	0.000	0.002
Civet SARSr-CoV civet010/Guangdong2004	0.000	0.002	0.003	0.001	0.001	0.001	0.004	0.002	0.000	0.000	0.003	0.001	0.003	0.001	0.003	0.003	0.005	0.003	0.005	0.005	0.006	0.003	0.000	0.003	0.002
Civet SARSr-CoV B039/Guangdong2004	0.000	0.003	0.004	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.002	0.001	0.003	0.001	0.000	0.001	0.006	0.008	0.005	0.006	0.003	0.000	0.003	0.003	0.003
Civet SARSr-CoV G209/GuangdongBlay003	0.000	0.002	0.003	0.002	0.000	0.001	0.000	0.002	0.000	0.000	0.002	0.001	0.003	0.001	0.000	0.002	0.005	0.011	0.005	0.006	0.003	0.000	0.003	0.004	0.003
Civet SARSr-CoV civet007/Guangdong2004	0.000	0.003	0.003	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.003	0.001	0.002	0.000	0.002	0.005	0.009	0.005	0.005	0.006	0.003	0.000	0.003	0.003	0.003
Civet SARSr-CoV A022/Guangdong2004	0.000	0.002	0.003	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.002	0.002	0.002	0.000	0.003	0.005	0.009	0.006	0.006	0.003	0.000	0.003	0.003	0.005	0.003
Civet SARSr-CoV civet209/Guangdong2004	0.000	0.004	0.004	0.001	0.000	0.001	0.004	0.002	0.000	0.000	0.002	0.002	0.003	0.000	0.001	0.006	0.008	0.005	0.006	0.007	0.000	0.003	0.003	0.005	0.003
Civet SARSr-CoV PC4_227/Guangdong2004	0.000	0.002	0.003	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.002	0.001	0.002	0.000	0.002	0.005	0.009	0.006	0.003	0.000	0.003	0.003	0.003	0.005	0.003
Human SARSr-CoV G20401/GuangzhouDec2003	0.000	0.003	0.004	0.001	0.000	0.001	0.004	0.002	0.000	0.000	0.003	0.002	0.002	0.000	0.002	0.009	0.008	0.005	0.010	0.007	0.000	0.003	0.003	0.003	0.003
Human SARSr-CoV PC4_136/Guangdong2004	0.000	0.002	0.004	0.001	0.000	0.001	0.004	0.002	0.000	0.000	0.002	0.003	0.003	0.000	0.001	0.006	0.008	0.005	0.006	0.003	0.000	0.003	0.003	0.004	0.003
Civet SARSr-CoV PC4_13/Guangdong2004	0.000	0.002	0.003	0.002	0.000	0.001	0.000	0.002	0.000	0.000	0.002	0.003	0.002	0.000	0.002	0.005	0.011	0.005	0.006	0.003	0.000	0.003	0.004	0.004	0.003
Civet SARSr-CoV A001	0.000	0.003	0.004	0.002	0.001	0.001	0.000	0.002	0.000	0.002	0.002	0.002	0.001	0.002	0.006	0.009	0.005	0.008	0.007	0.000	0.003	0.004	0.003	0.004	0.003
Human SARSr-CoV G20402	0.000	0.003	0.004	0.003	0.001	0.001	0.004	0.003	0.000	0.000	0.002	0.002	0.001	0.003	0.006	0.009	0.005	0.006	0.007	0.000	0.006	0.002	0.003	0.004	0.003
SARSr-Rs-BatCoV WIV16/R. sinicus/Yunnan/2013	0.023	0.033	0.045	0.028	0.019	0.013	0.037	0.020	0.024	0.024	0.019	0.042	0.040	0.040	0.067	0.066	0.079	0.058	0.028	0.014	1.064	0.009	0.031	0.015	0.035
SARSr-Rs-BatCoV Rs474/R. sinicus/Yunnan/2013	0.024	0.033	0.045	0.028	0.019	0.014	0.037	0.020	0.024	0.024	0.019	0.042	0.040	0.040	0.067	0.066	0.079	0.058	0.028	0.014	1.064	0.009	0.031	0.015	0.035
SARSr-Rs-BatCoV Rs4231/R. sinicus/Yunnan/2013	0.026	0.033	0.041	0.027	0.019	0.012	0.032	0.017	0.021	0.021	0.017	0.043	0.038	0.035	0.066	0.079	0.069	0.037	0.021	1.064	0.013	0.031	0.014	0.036	0.036
SARSr-Rs-BatCoV Rs4018/R. sinicus/Yunnan/2015	0.024	0.033	0.041	0.028	0.014	0.010	0.032	0.014	0.012	0.015	0.018	0.033	0.040	0.039	0.074	0.337	0.064	0.018	0.021	1.050	0.013	0.024	0.013	0.038	0.038
SARSr-Rs-BatCoV Rs7327/R. sinicus/Yunnan/2014	0.024	0.030	0.041	0.028	0.014	0.010	0.032	0.014	0.012	0.015	0.019	0.033	0.041	0.039	0.074	0.337	0.064	0.018	0.021	1.064	0.014	0.031	0.013	0.038	0.038
SARSr-Rs-BatCoV Rs3673/R. sinicus/Yunnan/2012	0.023	0.037	0.041	0.027	0.015	0.012	0.032	0.012	0.012	0.015	0.017	0.033	0.044	0.045	0.066	0.335	0.058	0.028	0.017	1.050	0.009	0.031	0.015	0.039	0.039
SARSr-Rs-BatCoV WIV1	0.023	0.036	0.044	0.026	0.014	0.009	0.032	0.012	0.012	0.015	0.017	0.033	0.043	0.043	0.069	0.337	0.069	0.027	0.013	1.050	0.013	0.031	0.015	0.041	0.041
SARSr-Rs-BatCoV RsSHC014/R. sinicus/Yunnan/2011	0.024	0.031	0.044	0.029	0.014	0.009	0.032	0.012	0.012	0.015	0.017	0.035	0.043	0.043	0.065	0.335	0.072	0.028	0.013	1.064	0.013	0.031	0.014	0.041	0.041
SARSr-Rs-BatCoV Rs4084/R. sinicus/Yunnan/2012	0.023	0.030	0.041	0.024	0.015	0.010	0.032	0.014	0.012	0.015	0.019	0.030	0.045	0.048	0.051	0.543	0.373	0.149	0.092	1.036	0.004	0.027	0.014	0.058	0.058
SARSr-Rs-BatCoV Rs4237/R. sinicus/Yunnan/2013	0.024	0.034	0.041	0.028	0.011	0.014	0.032	0.021	0.029	0.017	0.033	0.037	0.045	0.051	0.540	0.376	0.146	0.085	0.056	1.050	0.004	0.029	0.014	0.058	0.058
SARSr-Rs-BatCoV Rs4255/R. sinicus/Yunnan/2013	0.024	0.033	0.043	0.028	0.013	0.010	0.032	0.012	0.012	0.015	0.018	0.043	0.039	0.065	0.664	0.522	0.373	0.149	0.088	1.036	0.013	0.034	0.015	0.059	0.059
SARSr-Rs-BatCoV Rs4247/R. sinicus/Yunnan/2013	0.026	0.034	0.044	0.028	0.011	0.014	0.032	0.021	0.022	0.015	0.042	0.036	0.070	0.057	0.540	0.370	0.143	0.085	0.064	1.064	0.008	0.029	0.013	0.059	0.059
SARSr-Rs-BatCoV YN2013/R. sinicus/Yunnan/2013	0.032	0.041	0.046	0.027	0.014	0.009	0.032	0.017	0.015	0.017	0.018	0.031	0.042	0.064	0.664	0.483	0.382	0.153	0.089	0.921	0.013	0.023	0.063	0.063	0.063
SARSr-Rs-BatCoV Rs4081/R. sinicus/Yunnan/2012	0.021	0.033	0.044	0.026	0.018	0.013	0.037	0.019	0.024	0.022	0.018	0.043	0.041	0.067	0.545	0.525	0.382	0.149	0.089	1.050	0.000	0.034	0.018	0.060	0.060
SARSr-Rf-BatCoV Rf4092/R. ferromegalinum/Hubei/2005	0.024	0.040	0.042	0.028	0.018	0.008	0.020	0.019	0.008	0.024	0.020	0.027	0.042	0.073	0.079	0.602	0.382	0.159	0.095	0.024	1.064	0.034	0.017	0.062	0.062
SARSr-Rf-BatCoV Rf4092/R. ferromegalinum/Hubei/2005	0.023	0.039	0.043	0.038	0.025	0.012	0.029	0.020	0.027	0.027	0.021	0.042	0.039	0.070	0.070	0.522	0.389	0.145	0.091	1.036	0.004	0.027	0.023	0.062	0.062
SARSr-Rf-BatCoV YNLF_31C/R. ferromegalinum/Yunnan/2013	0.095	0.074	0.047	0.015	0.039	0.016	0.037	0.012	0.000	0.015	0.013	0.021	0.040	0.068	0.664	0.530	0.389	0.176	0.105	1.034	0.004	0.046	0.028	0.066	0.066
SARSr-Rf-BatCoV YNLF_34C/R. ferromegalinum/Yunnan/2013	0.090	0.073	0.048	0.015	0.039	0.016	0.037	0.012	0.000	0.023	0.014	0.021	0.040	0.067	0.665	0.530	0.389	0.176	0.105	1.034	0.009	0.044	0.028	0.067	0.067
SARSr-Rs-BatCoV Rp3/R. sinicus/Guangxi/2004	0.068	0.050	0.065	0.065	0.039	0.019	0.037	0.014	0.033	0.027	0.021	0.035	0.043	0.058	0.664	0.530	0.389	0.136	0.169	1.064	0.011	0.034	0.027	0.072	0.072
SARSr-Rs-BatCoV CZ013/R. sinicus/Guangxi/2013	0.068	0.050	0.070	0.063	0.039	0.023	0.033	0.015	0.030	0.022	0.025	0.045	0.045	0.064	0.664	0.530	0.389	0.136	0.169	1.064	0.011	0.034	0.027	0.072	0.072
UNVERIFIED: SARSr-Rp-BatCoV F46R.pusillus/Yunnan/2012	0.082	0.141	0.096	0.047	0.062	0.027	0.050	0.018	0.021	0.024	0.022	0.019	0.045	0.060	0.666	0.610	0.379	0.158	0.114	1.064	0.068	0.033	0.034	0.018	0.083
SARSr-Ra-BatCoV LYRa11/R. affinis/Yunnan/2011	0.080	0.142	0.096	0.101	0.082	0.075	0.103	0.072	0.068	0.062	0.032	0.062	0.086	0.092	0.101	0.361	0.100	0.142	0.114	1.090	0.018	0.056	0.029	0.092	0.092
SARSr-Cp-BatCoV Yunnan2011/C. haerophon pilosus/Yunnan/2011	0.084	0.141	0.054	0.101	0.082	0.072	0.094	0.063	0.071	0.060	0.053	0.080	0.082	0.051	0.112	0.532	0.373	0.151	0.098	1.064	0.013	0.056	0.029	0.093	0.093
SARSr																									



0.1

(caption on next page)

Fig. 4. Phylogenetic tree constructed based on the nt sequences of the RBD of the S protein of SARS-CoVs and SARSr-CoV. SARSr-Rf-BatCoVs are labeled with dots. Brackets in red, orange, yellow and blue represented the descending order of nt identity with respect to civet SARSr-CoV SZ16. The phylogenetic tree was constructed by Maximum Likelihood method with T92 + G as the substitution model by MEGA 6.0. The test of phylogeny was statistically supported by the bootstraps value calculated from 1000 trees. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

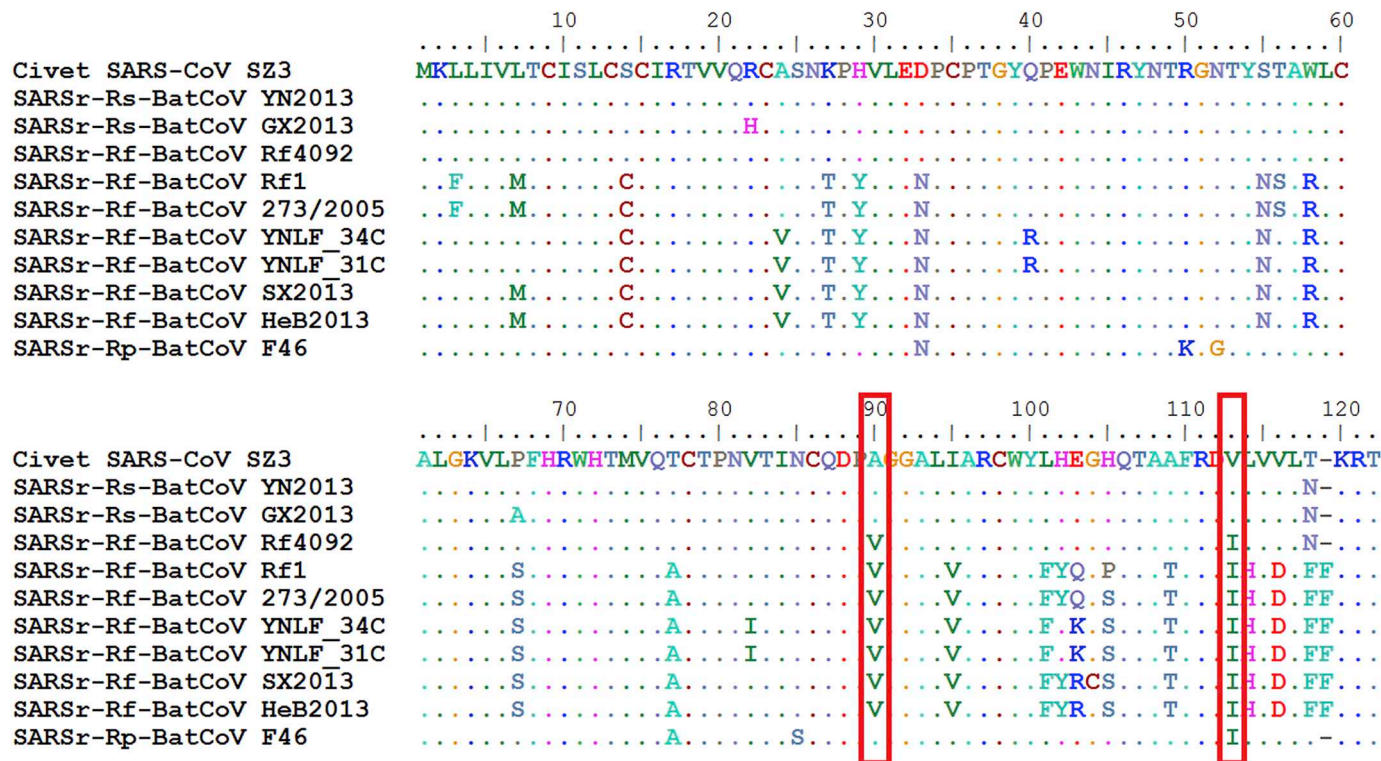


Fig. 5. Multiple alignment of type I ORF8 of SARSr-Rs-BatCoV YN2013, GX2013, civet SARSr-CoV and SARSr-Rf-BatCoV strain Rf4092 against the type II ORF8 from other SARSr-Rf-BatCoVs. Host-specific residues were highlighted with red boxes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Kunming. Nevertheless, the S proteins of the majority of bat SARSr-CoVs found in Chinese horseshoe bats in various part 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 question on how SARSr-BatCoVs 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.

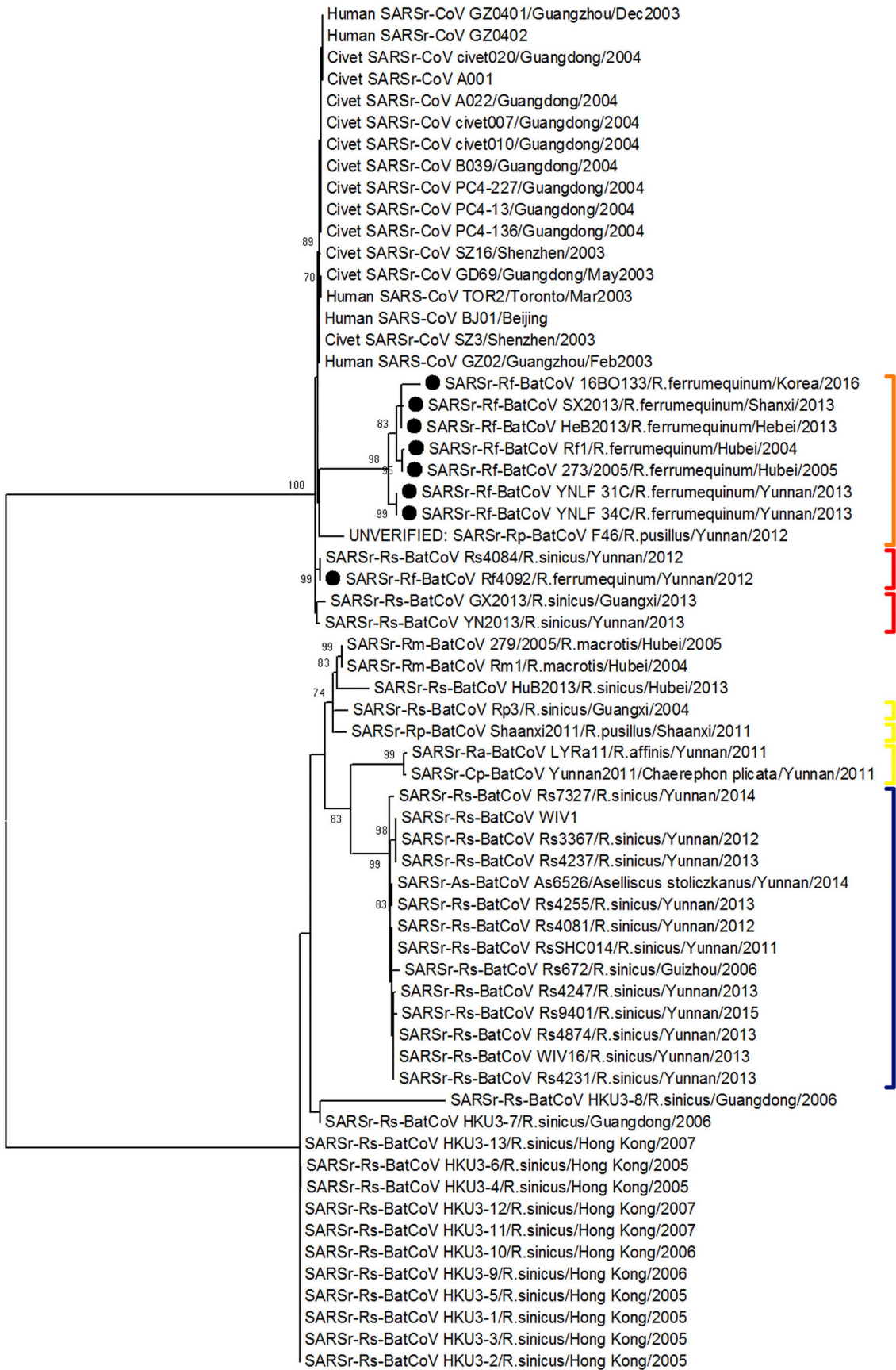
Interestingly, two novel SARSr-Rp-BatCoV strains ZXC21 and ZC45 isolated from least horseshoe bat in Zhoushan, Zhejiang province in eastern China were shown to be able to infect suckling rats, causing inflammation in the brain tissue and histological changes in the lung and intestine despite failed viral 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 Rf1. However, the mechanism of such infectivity in suckling rats was not investigated further, especially in light of the two retained deletion sites in the S genes of ZXC21 and ZC45, 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 and greater horseshoe bats. Detailed 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 SARSr-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). SARSr-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 SARSr-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 SARSr-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%



0.2

(caption on next page)

Fig. 6. Phylogenetic tree constructed based on the nt sequences of ORF8 of SARS-CoVs. SARSr-Rf-BatCoVs are labeled with dots. Brackets in red, orange, yellow and blue represented descending order of nt identity with respect to civet SARSr-CoV SZ16. The phylogenetic tree was constructed by Maximum Likelihood method with T92 + G as the substitution model by MEGA 6.0. The test of phylogeny was statistically supported by the bootstraps value calculated from 1000 trees. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

nt identity to the ORF8 of civet SARSr-CoV and have the greatest genetic distance observed over the whole genome of SARSr-CoV (Fig. 3B). They are detected across a wide range of bat species including *R. sinicus*, *R. macrotis*, *R. pusillus*, *R. affinis*, *A. stoliczkanus* 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 29 nt deletion. A similar 5-nt 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 SARSr-CoV with respect to civet SARSr-CoV. The importance of civet SARSr-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 SARSr-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 SARSr-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).

Referring 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 (NTD) 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, HKSAR Government; and University Development Fund, The University of Hong Kong.

References

- Campanella, J.J., Bitincka, L., Smalley, 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., Poon, L., Rottier, P.J., Talbot, P.J., Woo, P.C., Ziebuhr, J., 2011. Coronaviridae. 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.
- Drexler, J.F., Gloza-Rausch, F., Glende, J., Corman, V.M., Muth, D., Goettsche, M., Seebens, A., Niedrig, M., Pfefferle, S., Yordanov, S., Zhelyazkov, L., Hermanns, U., Vallo, P., Lukashev, 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.
- Ge, X.Y., Li, J.L., Yang, X.L., Chmura, A.A., Zhu, G., Epstein, J.H., Mazet, 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., Zhuang, Z.X., Cheung, 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.
- 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.
- 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, X.R., Zhang, Y.Z., Wang, N., Luo, D.S., Zheng, X.S., Wang, M.N., Daszak, P., Wang, L.F., Cui, J., Shi,

- Z.L., 2017. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* 13, e1006698.
- Hu, D., Zhu, C., Ai, L., He, T., Wang, Y., Ye, F., Yang, L., Ding, C., Zhu, X., Lv, R., Zhu, J., Hassan, B., Peng, Y., Tan, W., Wang, C., 2018. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg. Microbes Infect.* 7, 154.
- Hulswit, R.J., de Haan, C.A., Bosch, B.J., 2016. Coronavirus spike protein and tropism changes. *Adv. Virus Res.* 96, 29–57.
- Kan, B., Wang, M., Jing, H., Xu, H., Jiang, X., Yan, M., Liang, W., Zheng, H., Wan, K., Liu, Q., Cui, B., Xu, Y., Zhang, E., Wang, H., Ye, J., Li, G., Li, M., Cui, Z., Qi, X., Chen, K., Du, L., Gao, K., Zhao, Y.T., Zou, X.Z., Feng, Y.J., Gao, Y.F., Hai, R., Yu, D., Guan, Y., Xu, J., 2005. Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. *J. Virol.* 79, 11892–11900.
- Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Tsoi, H.W., Wong, B.H., Wong, S.S., Leung, S.Y., Chan, K.H., Yuen, K.Y., 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14040–14045.
- Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Wang, M., Lam, C.S., Xu, H., Guo, R., Chan, K.H., Zheng, B.J., Yuen, K.Y., 2007. Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. *Virology* 367, 428–439.
- Lau, S.K.P., Li, K.S.M., Huang, Y., Shek, C.T., Tse, H., Wang, M., Choi, G.K.Y., Xu, H., Lam, C.S.F., Guo, R., Chan, K.H., Zheng, B.J., Woo, P.C.Y., Yuen, K.Y., 2010a. Ecological and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J. Virol.* 84, 2808–2819.
- Lau, S.K.P., Poon, R.W.S., Wong, B.H.L., Wang, M., Huang, Y., Xu, H., Guo, R., Li, K.S.M., Gao, K., Chan, K.H., Zheng, B.J., Woo, P.C.Y., Yuen, K.Y., 2010b. Coexistence of different genotypes in the same bat and serological characterization of Roussetus bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. *J. Virol.* 84, 11385–11394.
- Lau, S.K.P., Lee, P., Tsang, A.K.L., Yip, C.C.Y., Tse, H., Lee, R.A., So, L.Y., Lau, Y.L., Chan, K.H., Woo, P.C.Y., Yuen, K.Y., 2011. Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. *J. Virol.* 85, 11325–11337.
- Lau, S.K., Woo, P.C., Yip, C.C., Fan, R.Y., Huang, Y., Wang, M., Guo, R., Lam, C.S., Tsang, A.K., Lai, K.K., Chan, K.H., Che, X.Y., Zheng, B.J., Yuen, K.Y., 2012a. Isolation and characterization of a novel Betacoronavirus subgroup a coronavirus, rabbit coronavirus HKU14, from domestic rabbits. *J. Virol.* 86, 5481–5496.
- Lau, S.K.P., Li, K.S.M., Tsang, A.K.L., Shek, C.T., Wang, M., Choi, G.K.Y., Guo, R., Wong, B.H.L., Poon, R.W.S., Lam, C.S.F., Wang, S.Y.H., Fan, R.Y.Y., Chan, K.H., Zheng, B.J., Woo, P.C.Y., Yuen, K.Y., 2012b. Recent transmission of a novel Alphacoronavirus, bat coronavirus HKU10, from Leschenault's Rousettes to Pomona leaf-nosed bats: first evidence of interspecies transmission of coronavirus between bats of different suborders. *J. Virol.* 86, 11906–11918.
- Lau, S.K., Li, K.S., Tsang, A.K., Lam, C.S., Ahmed, S., Chen, H., Chan, K.H., Woo, P.C., Yuen, K.Y., 2013. Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of Pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. *J. Virol.* 87, 8638–8650.
- Lau, S.K.P., Woo, P.C.Y., Li, K.S.M., Tsang, A.K.L., Fan, R.Y.Y., Luk, H.K.H., Cai, J.-P., Chan, K.-H., Zheng, B.-J., Wang, M., Yuen, K.-Y., 2014. Discovery of a novel coronavirus, China Rattus coronavirus HKU24, from Norway rats supports the murine origin of Betacoronavirus 1 and has implications for the ancestor of Betacoronavirus lineage A. *J. Virol.* 89, 3076–3092.
- Lau, S.K., Feng, Y., Chen, H., Luk, H.K., Yang, W.H., Li, K.S., Zhang, Y.Z., Huang, Y., Song, Z.Z., Chow, W.N., Fan, R.Y., Ahmed, S.S., Yeung, H.C., Lam, C.S., Cai, J.P., Wong, S.S., Chan, J.F., Yuen, K.Y., Zhang, H.L., Woo, P.C., 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.
- Lau, S.K., Tsang, A.K., Shakeel Ahmed, S., Mahbub Alam, M., Ahmed, Z., Wong, P.C., Yuen, K.Y., Woo, P.C., 2016. First genome sequences of buffalo coronavirus from water buffaloes in Bangladesh. *New Microbes New Infect.* 11, 54–56.
- Lau, S.K.P., Wong, E.Y.M., Tsang, C.C., Ahmed, S.S., Au-Yeung, R.K.H., Yuen, K.Y., Wernery, U., Woo, P.C.Y., 2018a. Discovery and sequence analysis of four Deltacoronaviruses from birds in the Middle East reveal interspecies jumping with recombination as a potential mechanism for avian-to-avian and avian-to-mammalian transmission. *J. Virol.* 92.
- Lau, S.K.P., Zhang, L., Luk, H.K.H., Xiong, L., Peng, X., Li, K.S.M., He, X., Zhao, P.S., Fan, R.Y.Y., Wong, A.C.P., Ahmed, S.S., Cai, J.P., Chan, J.F.W., Sun, Y., Jin, D., Chen, H., Lau, T.C.K., Kok, R.K.H., Li, W., Yuen, K.Y., Woo, P.C.Y., 2018b. Receptor usage of a novel bat lineage C betacoronavirus reveals evolution of MERS-related coronavirus spike proteins for human DPP4 binding. *J. Infect. Dis.* 218, 197–207.
- Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Cramer, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B.T., Zhang, S., Wang, L.F., 2005a. Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310, 676–679.
- Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.K., Huang, I.C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W.A., Guan, Y., Choe, H., Farzan, M., 2005b. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 24, 1634–1643.
- Liu, D.X., Fung, T.S., Chong, K.K., Shukla, A., Hilgenfeld, R., 2014. Accessory proteins of SARS-CoV and other coronaviruses. *Antivir. Res.* 109, 97–109.
- Menachery, V.D., Yount Jr., B.L., Debbink, K., Agnihothram, S., Gralinski, L.E., Plante, J.A., Graham, R.L., Scobey, T., Ge, X.Y., Donaldson, E.F., Randell, S.H., Lanzavecchia, A., Marasco, W.A., Shi, Z.L., Baric, R.S., 2015. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* 21, 1508–1513.
- Menachery, V.D., Yount Jr., B.L., Sims, A.C., Debbink, K., Agnihothram, S.S., Gralinski, L.E., Graham, R.L., Scobey, T., Plante, J.A., Royal, S.R., Swanstrom, J., Sheahan, T.P., Pickles, R.J., Corti, D., Randell, S.H., Lanzavecchia, A., Marasco, W.A., Baric, R.S., 2016. SARS-like WIV1-CoV poised for human emergence. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3048–3053.
- Muth, D., Corman, V.M., Roth, H., Binger, T., Dijkman, R., Gottula, L.T., Gloza-Rausch, F., Balboni, A., Battilani, M., Rihtaric, D., Toplak, I., Ameneiro, R.S., Pfeifer, A., Thiel, V., Drexler, J.F., Muller, M.A., Drosten, C., 2018. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci. Rep.* 8, 15177.
- Peiris, J.S.M., Lai, S.T., Poon, L.L.M., Guan, Y., Yam, L.Y.C., Lim, W., Nicholls, J., Yee, W.K.S., Yan, W.W., Cheung, M.T., Cheng, V.C.C., Chan, K.H., Tsang, D.N.C., Yung, R.W.H., Ng, T.K., Yuen, K.Y., 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361, 1319–1325.
- Song, H.D., Tu, C.C., Zhang, G.W., Wang, S.Y., Zheng, K., Lei, L.C., Chen, Q.X., Gao, Y.W., Zhou, H.Q., Xiang, H., Zheng, H.J., Chern, S.W.W., Cheng, F., Pan, C.M., Xuan, H., Chen, S.J., Luo, H.M., Zhou, D.H., Liu, Y.F., He, J.F., Qin, P.Z., Li, L.H., Ren, Y.Q., Liang, W.J., Yu, Y.D., Anderson, L., Wang, M., Xu, R.H., Wu, X.W., Zheng, H.Y., Chen, J.D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L.Y., Li, H., Chen, F., Di, B., He, L.J., Lin, J.Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X.J., Spiga, O., Guo, Z.M., Pan, H.Y., He, W.Z., Manuguerra, J.C., Fontanet, A., Danchin, A., Niccolai, N., Li, Y.X., Wu, C.I., Zhao, G.P., 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci.* 102, 2430–2435.
- Tang, X.C., Zhang, J.X., Zhang, S.Y., Wang, P., Fan, X.H., Li, L.F., Li, G., Dong, B.Q., Liu, W., Cheung, C.L., Xu, K.M., Song, W.J., Vijaykrishna, D., Poon, L.L., Peiris, J.S., Smith, G.J., Chen, H., Guan, Y., 2006. Phylogenetic and genetic diversity of coronaviruses in bats from China. *J. Virol.* 80, 7481–7490.
- Terada, Y., Matsui, N., Noguchi, K., Kuwata, R., Shimoda, H., Soma, T., Mochizuki, M., Maeda, K., 2014. Emergence of pathogenic coronaviruses in cats by homologous recombination between feline and canine coronaviruses. *PLoS One* 9, e106534.
- Tu, C., Cramer, G., Kong, X., Chen, J., Sun, Y., Yu, M., Xiang, H., Xia, X., Liu, S., Ren, T., Yu, Y., Eaton, B.T., Xuan, H., Wang, L.F., 2004. Antibodies to SARS coronavirus in civets. *Emerg. Infect. Dis.* 10, 2244–2248.
- Wang, N., Li, S.Y., Yang, X.L., Huang, H.M., Zhang, Y.J., Guo, H., Luo, C.M., Miller, M., Zhu, G., Chmura, A.A., Hagan, E., Zhou, J.H., Zhang, Y.Z., Wang, L.F., Daszak, P., Shi, Z.L., 2018. Serological evidence of bat SARS-related coronavirus infection in humans, China. *Virol. Sin.* 33, 104–107.
- Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W.K., Poon, L.L., Wong, S.S., Guan, Y., Peiris, J.S., Yuen, K.Y., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* 79, 884–895.
- Woo, P.C.Y., Lau, S.K.P., Li, K.S.M., Poon, R.W.S., Wong, B.H.L., Tsoi, H.-w., Yip, B.C.K., Huang, Y., Chan, K.-h., Yuen, K.-y., 2006a. Molecular diversity of coronaviruses in bats. *Virology* 351, 180–187.
- Woo, P.C.Y., Lau, S.K.P., Yip, C.C.Y., Huang, Y., Tsoi, H.W., Chan, K.H., Yuen, K.Y., 2006b. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. *J. Virol.* 80, 7136–7145.
- Woo, P.C., Wang, M., Lau, S.K., Xu, H., Poon, R.W., Guo, R., Wong, B.H., Gao, K., Tsoi, H.W., Huang, Y., Li, K.S., Lam, C.S., Chan, K.H., Zheng, B.J., Yuen, K.Y., 2007. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *J. Virol.* 81, 1574–1585.
- Woo, P.C., Lau, S.K., Lam, C.S., Lau, C.C., Tsang, A.K., Lau, J.H., Bai, R., Teng, J.L., Tsang, C.C., Wang, M., Zheng, B.J., Chan, K.H., Yuen, K.Y., 2012. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J. Virol.* 86, 3995–4008.
- Woo, P.C., Lau, S.K., Lam, C.S., Tsang, A.K., Hui, S.W., Fan, R.Y., Martelli, P., Yuen, K.Y., 2014a. Discovery of a novel bottlenose dolphin coronavirus reveals a distinct species of marine mammal coronavirus in Gammacoronavirus. *J. Virol.* 88, 1318–1331.
- Woo, P.C., Lau, S.K., Wernery, U., Wong, E.Y., Tsang, A.K., Johnson, B., Yip, C.C., Lau, C.C., Sivakumar, S., Cai, J.P., Fan, R.Y., Chan, K.H., Mareena, R., Yuen, K.Y., 2014b. Novel betacoronavirus in dromedaries of the Middle East, 2013. *Emerg. Infect. Dis.* 20, 560–572.
- Woo, P.C., Lau, S.K., Tsang, C.C., Lau, C.C., Wong, P.C., Chow, F.W., Fong, J.Y., Yuen, K.Y., 2017. Coronavirus HKU15 in respiratory tract of pigs and first discovery of coronavirus quasispecies in 5'-untranslated region. *Emerg. Microbes Infect.* 6, e53.
- World Health Organisation, 2004. Summary of Probable SARS Cases with Onset of Illness from 1 November 2002 to 31 July, pp. 2003.
- Wu, Z., Yang, L., Ren, X., Zhang, J., Yang, F., Zhang, S., Jin, Q., 2016. ORF8-related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J. Infect. Dis.* 213, 579–583.
- Yang, Z.-y., Werner, H.C., Kong, W.p., Leung, K., Traggiai, E., Lanzavecchia, A., Nabel, G.J., 2005. Evasion of antibody neutralization in emerging severe acute respiratory syndrome coronaviruses. *Proc. Natl. Acad. Sci.* 102, 797–801.
- Yang, L., Wu, Z., Ren, X., Yang, F., He, G., Zhang, J., Dong, J., Sun, L., Zhu, Y., Du, J., Zhang, S., Jin, Q., 2013. Novel SARS-like betacoronaviruses in bats, China, 2011. *Emerg. Infect. Dis.* 19, 989–991.
- Yang, X.L., Hu, B., Wang, B., Wang, M.N., Zhang, Q., Zhang, W., Wu, L.J., Ge, X.Y., Zhang, Y.Z., Daszak, P., Wang, L.F., Shi, Z.L., 2015. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. *J. Virol.* 90, 3253–3256.
- Zeng, L.P., Gao, Y.T., Ge, X.Y., Zhang, Q., Peng, C., Yang, X.L., Tan, B., Chen, J., Chmura, A.A., Daszak, P., Shi, Z.L., 2016. Bat severe acute respiratory syndrome-like coronavirus WIV1 encodes an extra accessory protein, ORFX, involved in modulation of the host immune response. *J. Virol.* 90, 6573–6582.