Geographic Evolution of Bat SARS-related Coronaviruses

A thesis submitted to

University of Chinese Academy of Sciences

in partial fulfillment of the requirement

for the degree of

Master of Natural Science

in Biochemistry and Molecular Biology

By

Yu Ping

Supervisor：Professor Cui Jie
Professor Shi Zheng-Li

Wuhan institute of Virology,Chinese Academy of Sciences

June 2019

University of Chinese Academy of Sciences
Declaration of Originality of Graduate Student Dissertation

I solemnly declare that the dissertation submitted by me is the result of my independent research work under the supervision of my supervisor. To the best of my knowledge, this dissertation does not contain any research results that have been published or written by any other individuals or collectives, except for the contents that have been cited in the text. Other individuals and groups who have contributed to the research covered in this thesis are clearly identified or acknowledged in the text.

Author's signature: xxxx
Date: 2019 5 30

Authorized Use Statement for Dissertation

I fully understand and agree to comply with the regulations of the Chinese Academy of Sciences on the preservation and use of the dissertation, that is, the Chinese Academy of Sciences has the right to retain copies of the dissertation sent to it, to allow the dissertation to be consulted, to publish all or part of the dissertation in accordance with the principles of openness of academic research and protection of intellectual property rights, and to preserve and compile the dissertation by photocopying, microfilming or other reproduction means. This statement applies to the dissertation that is confidential and delayed in disclosure after the declassification or delay period.

Author: XXX
Date: 2019 5 30

Abstract

Bats are the natural hosts of many emerging viruses, including many zoonotic viruses, and the outbreak of Severe acute respiratory syndrome coronavirus (SARS-CoV) in Guangdong in 2002 posed a serious health risk to humans. The discovery of SARS-like coronavirus (SARS-related CoV) in bats has demonstrated that bats are likely natural hosts for human SARS coronaviruses. Scientists have also identified bat SARS-like coronaviruses capable of utilizing the human ACE2 receptor as well as detected serologic evidence of human infection with such viruses, suggesting a potential future spillover risk to humans from these coronaviruses. Therefore, continuous surveillance of bat SARS-like coronaviruses is necessary.

Genetically diverse SARS-like coronaviruses have been identified in bats from different regions of China as well as Europe, Africa and Thailand since 2005, but their diversity, evolution and transmission are not well understood. Most of the analyses on SARS-like coronaviruses from non-Chinese countries are based on short sequences of coronavirus assays, and the evolutionary tree constructed based on this short sequence shows that SARS-like coronaviruses from China are more closely related to human SARS coronaviruses relative to other countries. To further reveal the evolution and transmission of SARS-like coronaviruses and their relationship with human SARS coronaviruses, this study performed large-scale detection and analysis of SARS-like coronaviruses from bat samples collected in twenty regions in China from 2011-2016 that were available in the laboratory. In this study, we amplified the full-length RdRp, S and ORF8 genes used for evolutionary analysis. Phylogenetic analysis revealed that the RdRp, S and ORF8 genes, the closest relatives of bat SARS-like coronavirus and human SARS coronavirus, were present in southern lineages (Yunnan, Guangxi, Guizhou, Guangdong and Chongqing), including Guangdong, where SARS coronavirus first appeared, suggesting that the ancestor of human SARS coronavirus originated from these southern regions. Sequence analysis revealed that the S gene of human SARS coronavirus probably originated from the SARS-like coronavirus in Yunnan, while the S gene was more geographically restricted than the ORF8 gene. In addition, the S gene of SARS-like coronavirus underwent frequent genetic recombination during the evolutionary process. Virus-host coevolution analysis revealed that SARS-like coronaviruses evolved mainly in homozygous bats of the Rhinolophus family, while host switching occurred mainly in different bat species of the Rhinolophus family. This study provides additional help to elucidate the geographic origin of SARS coronaviruses and the evolution of SARS-like coronaviruses in bats.

Keywords: bats, SARS, SARS-like coronavirus, geographic evolution, origin

Abstract

Bats are natural reservoirs for many emerging viruses, including many zoonotic viruses. In 2002, the severe acute respiratory syndrome coronavirus (SARS-CoV) broke out in Guangdong caused serious harm to human. The discovery of SARS-related coronavirus (SARSr-CoV) in bats demonstrated that bats are likely to be natural hosts for human SARS. CoV Scientists also discovered several SARSr-CoVs that can utilize the human receptor ACE2 and serological evidence of human infection, indicating that these coronaviruses have potential spillovers to humans in the future. Therefore, it is necessary to continue to monitor the bat SARSr-CoVS.

Since 2005, genetically diverse SARSr-CoVs have been found in different provinces of China and in bats in Europe, Africa and Thailand, but their diversity, evolution and spread have not been fully analysed. Non-Chinese countries analyze SARSr-CoVs mostly based on partial sequences for coronavirus detection, and the phylogenetic tree constructed showed that the SARSr-CoVs from China is closer to the human SARS-CoV than other countries. To further reveal the evolution and spread of SARSr-CoVs and their relationship with human SARS-CoV, this study conducted large-scale detection and analysis of bat SARSr-CoVs collected in twenty regions of China between 2011 and 2016. In this study, we amplified the full-length sequences of the RdRp, S and ORF8 genes for evolutionary analysis. Phylogenetic analysis showed that the RdRp, S and ORF8 genes of bat SARSr-CoVs closest to human SARS-CoV were from southern lineage (Yunnan, Guangxi, Guizhou, Guangdong and Chongqing), including the first occurence of SARS-CoV in Guangdong, indicating that the ancestor of human SARS-CoV originated in these southern regions. Sequence analysis of the ORF8 and S genes revealed that the S genes of the human SARS-CoV may have originated from the SARSr-CoVs in Yunnan, and the S gene is more geographically restricted than the ORF8 genes. In addition, the S genes of SARSr-CoVs had undergone frequent genetic recombination during their evolution. Co-evolution analysis of the viruses and their hosts showed that the SARSroCoVs evolved mainly in the same species of bats of the family Rhinolophidae, and the host shift mainly occurred in different bat species of the family Rhinolophidae. This study provides additional assistance for elucidating the geographic origin of the SARS-CoV and the evolution of the bat SARSr-CoVs.

Key words: Bats, SARS, SARSr-CoV, Geographical evolution, Origin

Table of Contents

Chapter 1 Introduction

1.1 Overview of Coronavirus

1.1.1 Classification of coronaviruses

Coronaviruses are a class of single-stranded positive-stranded RNA viruses with an envelope of about 80-120 nm in diameter, which have the second largest genetic material of all RNA viruses, and are called coronaviruses because of the presence of spines on the envelope, which are regularly arranged in a crown-like pattern on the surface of the capsule [1]. Infectious bronchitis virus (IBV), isolated from chickens in 1937, was the first coronavirus to be discovered, and more coronaviruses (HCoV-229E, HCoV-OC43, etc.) were subsequently discovered and reported [2-4]. For a long time, coronaviruses did not attract much attention and research on them was limited, mainly because they did not pose a major threat to human public safety and caused only mild symptoms of infection. It was not until the outbreak of Severe acute respiratory syndrome (SARS) between 2002 and 2003, which resulted in 8096 infections and 774 deaths, a mortality rate of approximately 10% [5-9], that the SARS outbreak attracted attention and attention to coronavirus-related research. The Middle east respiratory syndrome coronavirus MERS-CoV, a novel coronavirus outbreak in the Middle East in September 2012, is another highly pathogenic coronavirus that emerged after SARS-CoV. It is another highly pathogenic coronavirus after SARS-CoV [10].

According to the International Committee on Taxonomy of Viruses (ICTV), coronaviruses belong to the Nidovirales, Coronaviridae, subfamily Orthocoronavirinae [1, 11]. The subfamily Orthocoronaviridae can be further divided into four genera (genus). Alphacoronavirus, betacoronavirus, gammacoronavirus and deltacoronavirus. The genus alphacoronavirus can be evolutionarily divided into two families, A and B. The genus betacoronavirus can be evolutionarily divided into four families, A, B, C and D. The genera alphacoronavirus and betacoronavirus infect only mammals, while gammacoronavirus and deltacoronavirus infect mainly avian species, and a few can infect mammals [12]. The classification of coronaviruses found so far is shown in Figure 1.1. The genus alphacoronavirus includes human coronavirus (HCoV-229E, HCoV-NL63), bat coronavirus (CDPHE5, HKU10, HKU8, Mi-BatCoV 1A, Mi-BatCoV 1B, HKU2, Sc-BtCoV-512), porcine coronavirus (TGEV, PEDV, PRCV- ISU-1), feline coronavirus (FCoV), etc. The genus betacoronavirus includes human coronaviruses (SARS-CoV, MERS-CoV, HCoV-OC43, HCoV-HKU1), bat coronaviruses (SARSr-CoV, HKU4, HKU5, HKU9), porcine coronavirus (PHEV), and murine coronavirus (MHV). The genus gammacoronavirus includes avian coronavirus (IBV), beluga coronavirus (BWCoV). The genus deltacoronavirus includes avian coronaviruses (HKU11, HKU21, HKU13, HKU19, HKU12, HKU16, HKU20), and porcine coronavirus (HKU15) [13-16].

Figure 1.1 Phylogenetic tree of Coronaviruses [12].

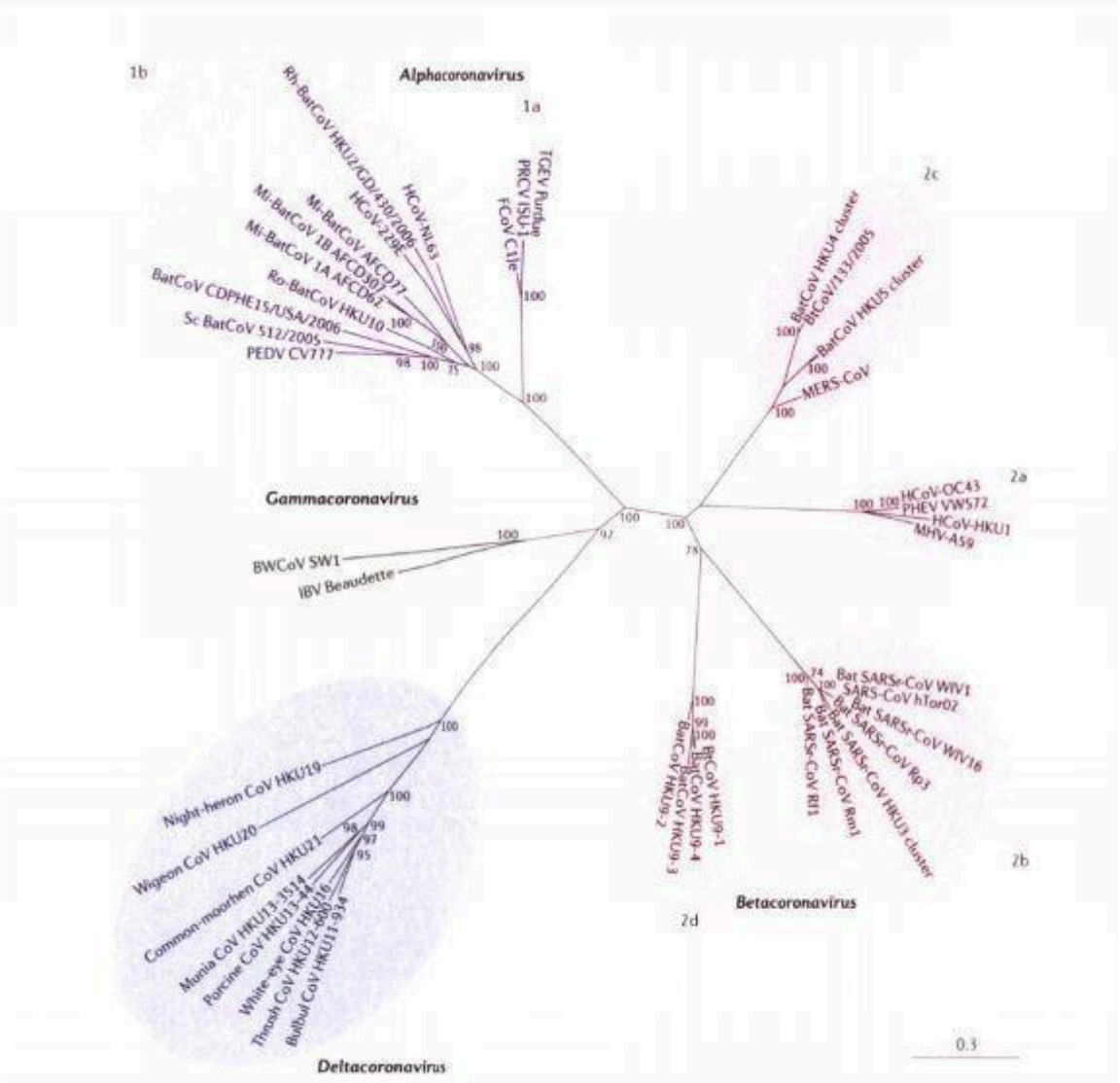


图 1.1 冠状病毒的系统发育进化树[12]。

Figure 1.1 Phylogenetic tree of Coronaviruses[12].

1.1.2 Epidemiology of coronaviruses

Coronaviruses are widely distributed and are found in several regions of the world such as China, United Kingdom, United States, Germany, Korea, Japan, and Thailand [17-19]. Coronaviruses are generally excreted through respiratory secretions and transmitted by oral fluids, air spray, and contact, usually infecting the epithelial cells of animals, with peak infection occurring in winter and spring, mainly causing respiratory infections with symptoms

including colds, fever, and vomiting, etc. The SARS coronavirus outbreak in 2002-2003 was transmitted mainly by respiratory and contact transmission, of which close droplet transmission was an important mode of transmission. The clinical symptoms caused by the outbreak of SARS in 2002-2003 were mainly fever, headache, malaise, diarrhea, dyspnea, etc. In addition, studies have shown that SARS coronavirus can survive in patients' excreta for up to 4 days, so the possibility of intestinal transmission cannot be ruled out [8,9]. 2012 outbreak of MERS virus was first identified in Saudi Arabia, causing symptoms similar to SARS coronavirus clinical symptoms, but its genome and receptors are different from SARS coronavirus, and most cases of MERS virus infection in Saudi Arabia are associated with exposure to dromedary camels, whereas most infections in other regions are associated with primary cases imported from the Middle East [20, 23].An outbreak of swine acute diarrhea syndrome coronavirus (Swine), which can cause death in pigs, in late October 2016 in Guangzhou pig farms The outbreak of Swine acute diarrhea syndrome coronavirus (SADS-CoV) in late October 2016 in a Guangzhou pig farm was characterized by severe acute diarrhea, vomiting, rapid weight loss, and mortality of up to 90% of piglets under 5 days of age, and genomic sequence analysis of SADS coronavirus showed a high degree of similarity with the genome sequence of bat coronavirus HKU2, suggesting that SADS coronavirus may have originated from bats, and serological investigation of pig farm workers in close contact with diseased pigs showed no evidence of further interspecies transmission of SADS coronavirus to humans [24].

1.1.3 Morphology and genome structure of coronaviruses

Coronaviruses are spherical or oval in shape and polymorphic in nature. There are three types of glycoproteins on the surface of the viral membrane: spike protein (S), envelope protein (E), and membrane protein (M). The S protein is a spherical, rod-shaped glycoprotein that extends 20 nm from the vesicle membrane and exists as a trimer in the coronavirus vesicle membrane, and is an important component in inducing cellular and humoral immunity, containing a host cell receptor binding recognition site and mediating cell membrane fusion; therefore, the S protein can determine the species heterogeneity and host range of viral infection [25,26]. The S1 subunit has a globular structure and is composed of the N-terminal domain (NTD) and the C-terminal domain (CTD), where the CTD, also known as the Receptor-binding domain (RBD), is critical for viral entry into the host cell. The S1 subunit is mainly involved in the recognition and binding of the virus to the host cell and is related to the cytophagy and virulence of the virus, while the S2 subunit constitutes the stem of the stinger protein and is mainly a helical structure, including a membrane fusion peptide (FP) region, two hydrophobic heptad repeat (Heptad) regions, and one heptad repeat (HR) region. The S2 subunit mainly mediates the fusion of the viral envelope and the cell membrane, and is also involved in the assembly and release of the virus [27-33]. The M protein, which spans the viral capsid, is the most abundant membrane penetrating protein on the viral envelope and host surface and has a highly conserved glycosylation sequence and is mainly responsible for viral particle assembly of viral particles and is also associated with the immunogenicity of viruses [35-37]. Some coronaviruses, such as MHV and PHEV, have another protein, Hemagglutinin-esterase (HE), in addition to these three proteins on the

capsid. HE protein exists as a dimer on the surface of viral particles, forming short protrusions with acetyl lipase activity, and plays an important role in virus It is not essential for viral replication, but seems to be important for host cell infection, and it shares 30% amino acid similarity with the hemagglutinin protein of influenza C virus, presumably obtained by genomic recombination with influenza C virus' [38-41]. The internal part of the viral particle is composed of viral RNA and the nucleocapsid protein (N), a helical structure with a diameter of 9-16 nm. The N protein is a nucleocapsid alkaline phosphorylated protein that binds tightly to the viral RNA to avoid damage to its genomic RNA [42-44]. The morphology of the coronavirus particle is shown in the upper right corner of Figure 1.2.

Coronavirus is currently the second largest RNA virus in terms of genome size, with a genome size range of 26.32 kb. Its RNA strand has a methylated cap structure at the 5' end and a PolyA tail structure at the 3' end, and this PolyA tail structure is very close to eukaryotic mRNA, which is also an important structural basis for its genomic RNA to play an important structural basis for the translation template role [45]. Both ends of the genome (5' and 3' ends) contain an Untranslated region (UTR), namely the 5'-UTR and 3'-UTR, respectively, where the 5'-UTR contains the ribosomal binding site and transcription initiation signal, and the 3'-UTR contains the transcription termination signal. The open reading frames shared by all coronaviruses from the 5' end to the 3' end of the genome are, in order, ORF1ab (pp1a and pp1b, both with a ribosomal frame shift, RFS), S (Spike), E (Envelope), M (Membrane) and N (Nucleocapsid), where ORF1ab is about 2/3 of the genome in size [46,47]. In addition to these shared open reading frames, there are some accessory open reading frames interspersed between structural protein genes in different genera of coronaviruses, such as ORF3 and ORF7 in bat coronavirus HKU2; ORF3, ORF6, ORF7, ORF8 and ORF9 in civet SARS coronavirus SZ3 (AY304486); ORF3 and ORF6 in infectious bronchitis IBV ORF3 and ORF6 of infectious bronchitis IBV, etc. The number and positions of the accessory genes of coronaviruses are variable on the genome, constituting the diversity of coronaviruses, but the relative positions of the common open reading frames are constant, ensuring the relative stability of the coronavirus structure. The genome structure of coronaviruses is shown on the left in Figure 1.2.

Figure 1.2 The genome organization and schematic representation of coronaviruses [12].

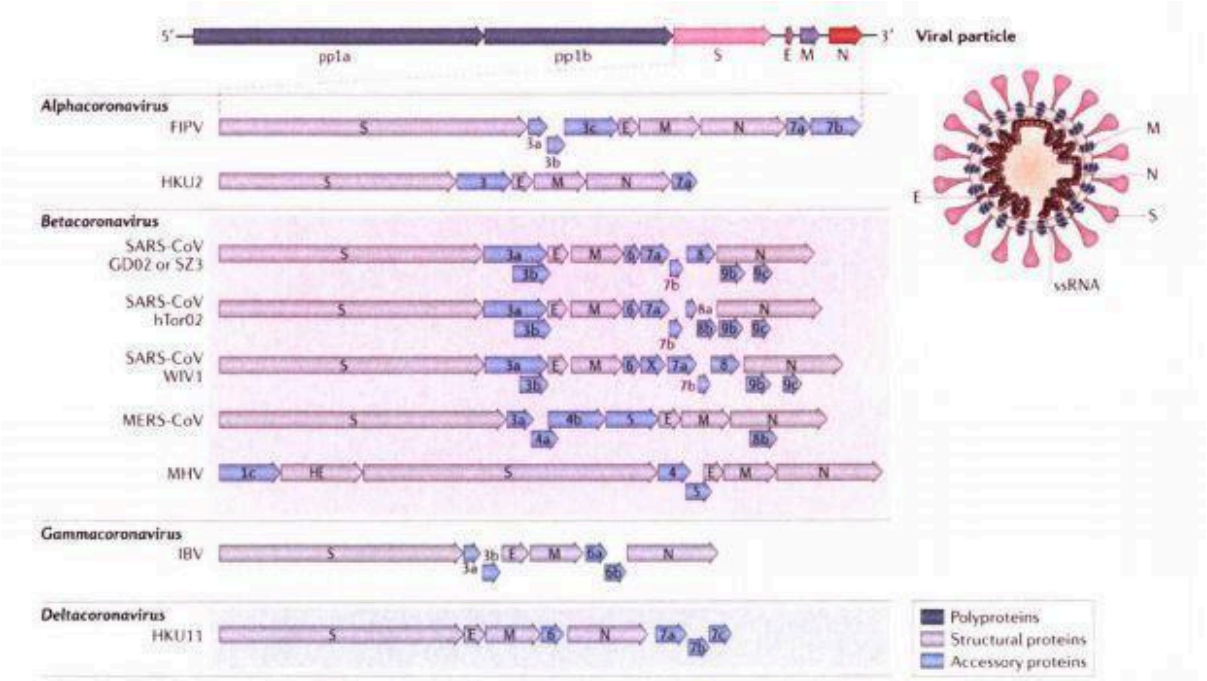


图 1.2 冠状病毒的基因组结构和结构示意图[12]。

Figure 1.2 The genome organization and schematic representation of coronaviruses[12].

1.1.4 Replication of coronavirus

After the coronavirus binds to specific receptors on the host cell surface, the viral capsid fuses with the cell membrane and then releases the viral nucleocapsid, which enters the cell and the genome is released into the cytoplasm, first using the viral genomic IWA as a translation template to express a variety of enzymes associated with viral replication, and then using the viral genomic RNA as a template to synthesize negative-stranded subgenomic RNA (Subgenomic RNA), and then use this as a template to synthesize genomic RNA, as well as transcription from its 3' end until it encounters the transcription regulatory sequence (TRS) upstream of the open reading frame, and by base-pairing, the nascent The transcription continues through the 5' end of the genome, resulting in Subgenomic mRNA, which is then used as a template to translate the structural proteins required for the virus, and the structural proteins and genomic RNA are transferred to the endoplasmic reticulum to assemble the new transcript. to the endoplasmic reticulum to assemble new coronavirus particles and secrete the virus out of the body via the Golgi apparatus [45,48,49]. Each mRNA has a leader sequence from the 5' end of the genomic RNA, and coronavirus constituents [49].

Figure 1.3 The genome replication of coronavirus [49].

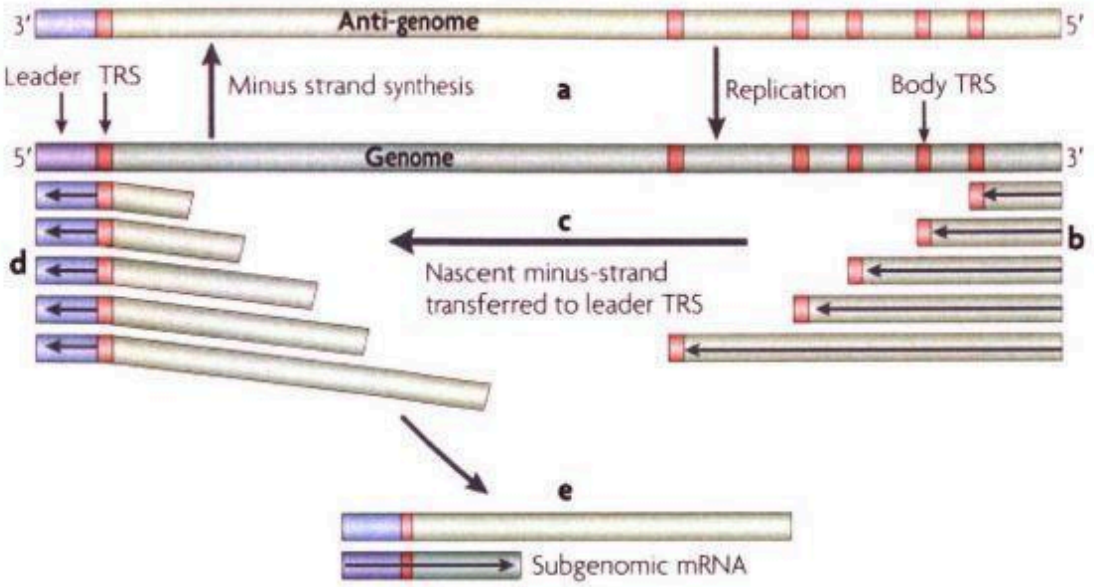


图 1.3 冠状病毒的基因组复制[49]。

Figure 1.3 The genome replication of coronavirus[49].

1.1.5 Genetic recombination of coronaviruses

Due to the large genome of coronaviruses, which may accumulate a large number of errors during replication, and the absence of post-transcriptional modification shearing processes during translation of structural proteins, viruses need to survive and must have a genetic mechanism to minimize this damage, and viral genetic recombination is one mechanism. Coronaviruses use a unique set of transcripts that produce a series of subgenomic RNAs during viral replication, which may lead to template strand switching, and jumps between different template strands increase the homologous recombination of coronavirus or other virus-related genes from different lineages, in addition, the circulation and recycling of coronaviruses in different hosts may accelerate the rate of recombination, and recombination makes coronaviruses plastic and may generate new recombinant strain [15,50,51]. Genetic recombination has been identified in animal and human coronaviruses such as MHV, TGEV, FCoV, CCoV, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, and MERS-CoV [15]. Recombinant strains containing a single species of non-nodal RNA virus were obtained by infecting astrocytic glioma cells (DBT cells) with two mutant strains (A59 and JHM) of mouse hepatitis virus (MHV) at non-permissive temperatures, and sequence analysis showed that the genome of this recombinant strain was mainly from A59, with 3 kb at the 5' end from JHM, indicating that the leading sequence (leader sequence) of JHM could be fused into the mRNA of different strains during RNA transcription. The recombinant strain was divided into

five genotypes (A-E), and the recombination analysis showed that the D genotype of HCoV-OC43 was fused to the mRNA of different strains through the B and C genotypes of HCoV-OC43 in the nonstructrual protein (Nsp) [52].Recombinant strains were obtained by recombination of Nsp2-Nsp3 and Nsp12 [53]. recombination site analysis of HCoV-NL63 showed that recombination of HCoV-NL63 occurred in S1 and S2 of the S gene [54]. Most recombination events mainly involve homologous recombination of related viruses, and genetic recombination can occur not only in homologous coronaviruses, but also in different species of coronaviruses, or even in different families. Using pan-coronavirus priming techniques and deep sequencing, scientists identified a novel coronavirus, Ro-Bat coronavirus GCCDC1 (Ro-BatCoV GCCDC1), from a bat rectal swab sample collected in Xishuangbanna, Yunnan Province, which is similar in genome structure and composition to Ro-Bat coronavirus (Ro-BatCoV HKU9), had a gene at the 3' end of its genome that was not found to be homologous in known coronaviruses. Subsequent sequence and evolutionary analyses showed that this gene is closely related to the p10 gene of an ancient bat orthorectal virus, and subgene mRNA and cellular level experiments further indicated that both are functional genes capable of inducing cellular syncytium synthesis, which is the first time that cross-family heterologous recombination has been identified between a single-stranded positive-stranded RNA virus and a double-stranded segmented RNA virus [55].

1.1.6 Cross-species transmission of coronaviruses

Coronaviruses are single-stranded RNA viruses that are susceptible to mutation, which can increase the species diversity of coronaviruses and also potentially confer rapid adaptation to new hosts for cross-species transmission to occur. Cross-species transmission of viruses is the transfer of viruses from one host to another and involves processes that include contact between the virus and a different host, infection and replication of the initial individual, and effective transmission in the new host population [56]. Viral entry and host innate immunity are the primary hurdles to overcome for viral host switching. The host cell receptor for MERS coronavirus, Dipeptidyl peptidase 4 (DPP4), is well conserved between bats and humans, and antibodies against DPP4 can block viral infection [57,58]. The mutation rate of coronaviruses is moderate compared to other single-stranded RNA viruses, with an average substitution rate of approximately 10'-4/year per locus [54]. The base mutation rate of IBV in the highly variable region of the S gene is 0.003-0.006/year, and the base mutation rate of HCoV-229E in the S gene region is approximately 3.0 × 10’-4/year [54, 59]. In vitro experiments have demonstrated that the probability of mutation per base during a round of genome replication of SARS coronavirus is about 10'-6 [50]. The accumulation of mutations, especially in the receptor-binding region (RBD), may lead to altered viral recognition of the receptor, resulting in cross-host infection. The crystal structure of the RBD of the SARS coronavirus S gene and the human ACE2 binding shows that the RBD of SARS coronavirus contains a core structure and a Receptor-binding motif (RBM), and the variations in amino acids at two key sites (479 and 487) of the RBM determines the higher affinity of SARS coronavirus RBD with human ACE2 and lower affinity with civet ACE2, and the alteration of amino acids at these two key sites may allow SARS coronavirus to cross the

species barrier between civet and human [12]. The deletion of more than 600 nucleotides in the S1 gene of the stinging protein alters the tropism of the virus to the host cell, transforming porcine transmissible gastroenteritis virus (TGEV) into porcine respiratory coronavirus (PRCV) [18]. Genetic recombination of coronaviruses can also lead to cross-species transmission of the virus. If recombination occurs in the spike protein gene, especially in the RBD region, it may result in altered cellular recognition properties of the recombinant virus, which in turn may cause cross-host transmission. Sequence analysis of the variable region of the S protein gene of six canine coronavirus strains revealed that the 5' end of the S gene of five strains was highly homologous to the 5' end of the S gene of feline infectious peritonitis virus (FIPV). One isolate, CCoV UCD-1, however, had high homology with the 5' end of the S gene of TGEV, while this isolate was found to differ from other isolates in its ability to infect porcine testicular cells, suggesting that RNA recombination occurs between antigenically related coronaviruses [60]. Analysis of infectious bronchitis IBV PP14 isolated from chickens revealed the presence of homologous recombination between IBV Mass 41 and Ark 99 at the 5' end of the S1 gene, which alters serotype-specific epitopes [61,62]. Natural hosts carrying viruses, through natural or artificial contact, may lead not only to co-infection of viruses from different host sources, but also to recombination between different viruses in co-infected hosts, providing the possibility of cross-species transmission. The large RNA genome in coronaviruses allows for additional plasticity in genomic modifications through mutation and recombination, increasing the potential for intraspecific variation, host transformation, and the generation of new coronaviruses [15].

1.2 Overview of SARS-like coronaviruses

1.2.1 Emergence and traceability of SARS coronavirus

In November 2002, the first case of Severe acute respiratory syndrome (SARS), also known as "atypical pneumonia" or "SARS," emerged in Foshan, Guangdong, China. "Subsequently, the SARS epidemic spread to Hong Kong and other parts of China, and rapidly spread to more than 20 other countries, including Vietnam and Canada, until the epidemic was gradually eliminated in mid-2003 [5]. On April 16, 2003, the World Health Organization (WHO) announced that the cause of acute respiratory syndrome (ARS) was a novel coronavirus, named SARS-CoV. As of July 2003, according to WHO statistics, SARS coronavirus caused infections in 29 countries, including China and Singapore, resulting in 8096 infections, of which 774 deaths. Between December 2003 and January 2004, there were five other cases of infection due to zoonotic diseases, but no deaths or human-to-human transmission occurred.

SARS coronavirus particles are mostly spherical or oval in shape, occasionally irregular in shape, with a diameter of about 60-120 nm, an envelope, and radial spine-like protrusions on the surface.The genome structure of SARS coronavirus is similar to that of other coronaviruses, including 16 non-structural proteins Nsp 1-Nsp 16 in ORF1ab and structural

proteins (S protein, E protein, M protein, N proteins and other accessory proteins) [45, 63]. SARS coronavirus infects respiratory epithelial cells mainly by droplet and aerosol transmission, resulting in clinical symptoms such as fever, headache, malaise, diarrhea, and respiratory distress.

After the SARS coronavirus outbreak, researchers conducted an epidemiological investigation. The epidemiological findings revealed that all of the original cases had been in contact with animals, indicating that SARS coronavirus may have originated from animals. Subsequently, animal tracing was conducted in a wild market in Shenzhen, Guangdong, and antibodies to SARS coronavirus were found in civets, raccoons and badgers. Whole-genome sequencing analysis of SARS coronavirus isolated from civet specimens revealed that this SARS coronavirus and the SARS coronavirus isolated from humans have more than 99% homology, and the results of the study indicated that SARS coronavirus may originate from animals such as civets and infect humans through cross-species transmission [64,66]. To further verify whether civets are the natural hosts of SARS coronavirus, researchers tested civets from wildlife farms for the virus, however, no positive SARS coronavirus was detected from them [67,68]. Artificial infection of civets with SARS coronavirus produced the same symptoms as SARS coronavirus infection in humans. These results suggest that civets are not natural hosts of SARS coronavirus, but play the role of intermediate hosts in the transmission of SARS coronavirus to humans.

1.2.2 Discovery of SARS-like coronavirus

In 2005, a SARS-like coronavirus, called SARS-like coronavirus (SARS-related CoV, SARS-like CoV, SARSr-CoV or SL-CoV), was found simultaneously in Chinese chrysanthemum-headed bats by a research group at the University of Hong Kong and the laboratory of Ms. Zhengli Shi, and the finding provided a traceability to bats as a possible natural host of SARS coronaviruses [69,70]. Subsequently, more bat SARS-like coronaviruses have been reported in Asia, Europe, and Africa [71-81]. In particular, some SARS-like coronaviruses, such as WIV1 (KF367457), were able to use angiotensin converting enzyme II (ACE2) from civets, bats, and humans as receptors for virus entry into cells, a finding that further confirmed the origin of human SARS coronaviruses in bats and also suggests that these SARS-like coronaviruses can directly infect humans without the need for other intermediate hosts [82]. In addition, serological evidence of human SARS-like coronavirus infection detected by ELISA near bat roosts in Yunnan, China, indicates the potential for spillover of SARS-like coronaviruses from bats to humans [83].

1.2.3 Genetic diversity of SARS-like coronaviruses

Both SARS coronaviruses and SARS-like coronaviruses belong taxonomically to the coronaviridae, the betacoronaviridae family B viruses. Their genome structure is similar to that of other coronaviruses, consisting of a non-structural protein gene that accounts for 2/3 of the genome size and several downstream structural protein (S protein, E protein, M protein, N protein and other accessory proteins) genes [84]. By comparing the genomes of SARS coronaviruses and SARS-like coronaviruses, the main differences were found in nonstructural protein 3 (Nsp 3), ORF3, S, and ORF8, with S and ORF8 showing the greatest differences [68,85,86]. A bat SARS-like coronavirus strain Rs672/2006 (FJ588686) found in Guizhou, China, has a 579-nt deletion in the nsp3 region, and similar deletions have been found in late epidemic human SARS coronaviruses such as ShanghaiQXC2 (AY463060), and such deletions in Rs672/2006 and human SARS coronaviruses may occur independently [76].The SARS coronavirus genome encodes a 154 amino acid sequence ORF3b in the second half of the internal part of the ORF3a gene, and most of the coding sequence of ORF3b is shared with ORF3a, except for the C-terminus, where ORF3b encodes an interferon antagonist, and bat SARS-like coronaviruses and SARS coronaviruses have high amino acid sequence similarity in the ORF3a region, which can reach more than 96%, while ORF3b is variable in size in bat SARS-like coronaviruses (54-154 amino acids), different truncations exist at the C-terminus of ORF3b, and different ORF3b proteins exhibit different interferon antagonist activities, in addition, some bat SARS-like coronaviruses The main difference between SARS coronavirus and SARS-like coronavirus in the S gene is expressed in the receptor binding motif (RBM) region, and the amino acids in the RBM region of bat SARS-like coronavirus were compared with those of the human or civet SARS coronavirus. The majority of bat SARS-like coronaviruses had two deletions in this region (5 amino acids and 12 or 13 amino acids), while the Bulgarian strain BM48-31 (GU190215) from Europe had only one deletion in this region, and a few SARS-like coronaviruses such as RsSHC014 (KC881005), Rs3367 (KC881005), etc. have RBM amino acid sequences as long as SARS coronaviruses, and these bat SARS-like coronaviruses without deletions in the RBM region were shown to be able to use civet and human ACE2 as receptors for virus entry into host cells, while those bat SARS-like coronaviruses with deletions in this region are not able to use civet and human ACE2 to enter host cells [73,79,89,90]. The ORF8 gene is an accessory protein gene that is highly variable during various periods of the SARS coronavirus epidemic. In the early stages of the SARS coronavirus outbreak, most SARS coronaviruses have an intact single gene for ORF8, and a few have an 82nt deletion in ORF8; while in the middle and late , the majority of SARS coronaviruses have a 29nt deletion in ORF8, causing ORF8 to cleave into two small open reading frames, ORF8a and ORF8b. ORF8a protein can increase SARS coronavirus replication and induce caspase-dependent apoptosis, ORF8b can inhibit the interferon signaling pathway, and there are a few SARS coronaviruses in the late stage due to 415nt deletion resulted in complete absence of ORF8, indicating that ORF8 is not essential for SARS coronavirus replication [91-94]. Deletion of ORF8 has also been reported in bat SARS-like coronaviruses, and HKU3-8 (GQ153543) isolated from a Guangdong bat sample was missing 26nt in the ORF8 region, causing its ORF8 to be divided into ORF8a, ORF8b and ORF8c, and the ORF8 of Rs4084 (KY417144) isolated from Yunnan bat samples was divided into ORF8a and ORF8b due to a 5 nt deletion, while the ORF8 of BM48-31 isolated from the European Bulgarian daisy-headed bat (Rhinolophus blasii) was completely missing. Overall, the majority of bat SARS-like coronaviruses reported so far have no deletion of ORF8 [72,73,87,95,96]. In addition to this, some bat SARS-like coronaviruses such as WIV1 and WIV16 have an additional novel gene between genomic

ORF6 and ORF7, called ORFX, which is involved in regulating the host immune response [87,89,97].

1.2.4 Epidemiology of SARS-like coronaviruses

SARS-like coronaviruses are mainly found in the family Rhinolophidae, including Rhinolophus affinis, Rhinolophus blasii, Rhinolophus ferrumequinum, Rhinolophus hipposideros, Rhinolophus macrotis, Rhinolophus monoceros, Rhinolophus pearsonii, Rhinolophus pusillus, Rhinolophus rex, Rhinolophus sinicus, Rhinolophus thomasi, and a few other species, including Aselliscus stoliczkanus, Hipposideros commersoni, Hipposideros larvatus, etc. [4] The SARS-like coronavirus has been detected in more than a dozen countries, including China, Kentucky, and the United States. In 2013, researchers identified and isolated bat SARS-like coronavirus WIV1 capable of utilizing the ACE2 receptor, further demonstrating the bat origin of SARS coronavirus [82]. In addition, full-length and chimeric WIV1-CoV constructed by reverse genetics systems were efficiently replicated in human airway cultures and in vivo experiments, indicating that the bat SARS-like coronavirus WIV1-CoV cluster has the ability to be transmitted from bats to humans, while the attenuation of mouse pathogenesis in WIV1-CoV spinosomal protein-mediated infection relative to SARS-CoV was also observed to fail be ameliorated by enhancing the binding of the spike-in protein to the host receptor, suggesting that there may be other influences besides the key receptor binding region, RBD, including adaptations outside the spike-in protein and changes in the spike-in protein independent of the receptor binding region to explain this attenuation, and this requires further corroboration in non-human primates [90]. Furthermore, comparing sera from humans with little access to bats with sera from humans living near bat caves with circulating SARS-like coronaviruses, the possibility of human infection by bat SARS-like coronaviruses or other related viruses was confirmed by ELISA assays that detected positivity in sera from the latter, however, subsequent virus neutralization tests (Neutralization test) and protein blotting (Western blot) assays failed to detect antibodies to SARS-like coronavirus RBD, possibly due to the immune response and the high diversity of bat SARS-like coronavirus spike proteins [83]. Although the inhabitants of this site do not exhibit severe symptoms of infection, those bat SARS-like coronaviruses that can utilize the human ACE2 receptor are still a threat of spillover infection to humans in the future; therefore, continued detection and investigation of bat SARS-like coronaviruses is necessary.

1.2.5 Recombination of SARS-like coronavirus

Recombination is common in coronaviruses and plays an important role in the evolution of viruses, by which new viruses can be generated and the host range of the virus can be extended. Recombination events have also been found in SARS coronaviruses and SARS-like coronaviruses. Recombination analysis of SARS coronaviruses and bat

SARS-like coronaviruses revealed the presence of two important recombination hotspots, the S gene and ORF8 gene, possibly due to the high variability of these two genes, with recombination of the S gene occurring mainly between NTD and RBD [87]. When using the SARS coronavirus SZ3 from civet as a query sequence, potential recombination events were identified in the spacer regions of the Nsp16 and S genes of SARS-like coronaviruses Rm1 (DQ412043), Rf1 (DQ412042) and Rp3 (DQ071615) [73]. After 5 years of monitoring bats in a cave in Yunnan, a natural gene pool of SARS coronaviruses was identified, and all genomic components of SARS coronaviruses could be found in this cave, and recombination analysis of 15 full-length SARS-like coronaviruses as well as the genome of SARS coronaviruses showed that regions within the S gene and ORF8 of SARS-like coronaviruses underwent frequent recombination, and by further recombination analysis, it was found that WIV16 could be obtained by recombination of WIV1, Rs4231 (KY417146) and Rs4081 (KC880999), Rs4084 could be obtained by recombination of RsSHC014 and Rf4092 (KY417145), and WIV16 and Rf4092 could recombine to obtain SZ3, and it is speculated that the direct ancestor of SARS coronavirus may originate from recombination between these ancestral strains of bat SARS-like coronavirus in Yunnan [87].

1.3 Bats and emerging viruses

1.3.1 Bats

Bats are the second largest group of mammals after rodents and are the only group of mammals with the ability to fly, with over 1200 species in 19 families, accounting for about 20% of all mammalian species [98]. Taxonomically, bats belong to the Animalia, Chordata, Mammalia, and Chiroptera, which can be further divided into two suborders, Megachiroptera and Microchiroptera. Microchiroptera), also known as fruit-eating bats and insect-eating bats, the former are large, have a keen sense of smell, feed on fruits, rely mainly on vision to identify objects, and include only the family Pteropodidae (fox bats), which are distributed in tropical and subtropical regions of the Old World, while the latter are small, usually have degenerate vision, rely mainly on sonar to identify objects, and feed mainly on insects. There is a wide variety of species with a wide distribution [99]. A few bats can also feed on animal blood or fish. Bats are long-lived, and some bat populations roost together, while others roost separately. Bats generally inhabit caves, cracks in old buildings, ceilings, and tree caves. During winter, many bats roosting in the woods migrate to warmer areas, over distances of up to thousands of kilometers, where their metabolic capacity is reduced during hibernation. With the exception of the polar regions and some islands in the oceans, bats are found worldwide, mostly in tropical and subtropical regions. Most bats rest during the day and forage at night. In addition, bats, like insects such as bees, play an important role in the ecosystem and are important pollinators.

1.3.2 Bats and emerging viruses

More than 200 viruses from 27 families and 37 genera have been detected from 12 families of bats, including Hantavirus, Rabies virus, Coronavirus, Ebolavirus (EBoV), Hendra virus (Hendravirus), Nipahvirus, Marburg virus, etc. [100-102]. Hantaviruses belong to the Bunyaviridae family and are enveloped segmented negative-stranded RNA viruses that cause two main symptoms: Hantavirus pulmonary syndrome (HPS) and Hemorrhagic fever with renal syndrome (HFRS). The former is prevalent mainly in the United States, with cases also found in Argentina, Brazil, Paraguay, Germany, and other countries, and the latter is prevalent mainly in Asian countries such as China, Japan, and North Korea [103]; rabies virus belongs to the genus Rhabdoviridae (Lyssavirus), which is transmitted through animal bites and can cause severe acute infectious diseases of the central nervous system with an incubation period and high lethality [104]; Ebola virus, which was first discovered in 1976 in the Ebola River region of southern Sudan and the Democratic Republic of the Congo, belongs to the Filoviridae family and causes Ebola in humans and other primates. The virulent infectious disease of hemorrhagic fever, which is the most lethal viral hemorrhagic fever in the world today, with a mortality rate between 50-90%, has some spatial and temporal limitations, presenting endemic epidemics, confined to the tropical rainforests of Central Africa and the savannas of Southeast Africa, and imported cases of Ebola virus infection have been found in the United States, the United Kingdom, Switzerland, and other countries [105-107]. Animal tracing experiments of viruses have confirmed that bats are natural hosts of several viruses, such as SARS coronavirus, Hendra virus, rabies virus, and Ebola virus [99]. The ability of bats to fly requires the digestion of large amounts of energy, and in the process of digesting large amounts of energy, large amounts of waste are produced. Bats have evolved repair mechanisms that can adapt to high levels of DNA damage, thus avoiding the loss of their own DNA from metabolic waste. Secondly, bats have a stronger intrinsic immune response system at background levels compared to other animals, such as a high background level of type I interferon IFN-α, this high level of background intrinsic immune system causes the bat itself to maintain a high level of alertness to foreign infection, unlike other mammals that require induction to express immune molecules, and a faster antiviral response: then, one of the antiviral immune channels in the bat, the "interferon gene stimulating protein I" channel, is inhibited and bats are able to develop tolerance to the virus without exhibiting morbidity or eliciting a strong immune response [108-110]. In short, the antiviral immunity and immune tolerance of bats are balanced during the process of virus infection, allowing bats to carry the virus without causing disease. In addition, the ability of viruses carried by bats to spread across species to humans is also associated with many factors, firstly, the wide distribution of bats, which are distributed worldwide except for polar regions and some islands in the oceans, and the high chance of contact between humans and bats, where long-term contact may cause changes in the virus to adapt to new hosts, such as the spread of SARS coronavirus from civets to humans, which involves amino acid changes at key loci in the RBD region changes, followed by the fact that bats and humans are both mammals, not very distantly related evolutionarily, and some signaling pathways and receptors are similar, providing conditions for cross-species transmission of viruses; in addition, the ability of bats to fly can provide more opportunities for different populations to communicate, which can facilitate the exchange of

viruses they carry and cause genetic recombination of different viruses, etc., also providing conditions for cross-species transmission of viruses [12 ,111].

1.4 Research content and significance of this thesis

The SARS coronavirus outbreak between 2002 and 2003 posed a serious risk to human health and caused widespread concern worldwide. Animal tracing proved that civets only played an intermediate host role in the virus transmission chain, and the SARS-like coronavirus found in Chinese daisy-headed bats in 2005 suggested that bats might be the natural host of SARS coronavirus. The discovery and isolation of SARS-like coronavirus WIV1, which can utilize the ACE2 receptor, further confirmed the origin of SARS coronavirus in bats in 2013. Subsequently, more viruses have been found in bats, including rabies virus, hantavirus, and Ebola virus, and therefore, long-term studies of these bat-borne viruses are necessary to assess the risk of transmission or cross-species transmission and subsequent prevention and control.

Although SARS coronaviruses were gradually controlled and eliminated after late 2003 through a series of measures (including control of the source of infection and drug treatment), those SARS-like coronaviruses that can exploit the human ACE2 receptor remain a potential risk of spillover infection in the future. Since 2005, bat SARS-like coronaviruses have been reported in China, Kenya, Nigeria, Ghana, Thailand, Bulgaria, France, Spain, Luxembourg, Korea, Slovenia, and Italy. The origin of SARS coronaviruses in bats has been confirmed, and previous evolutionary analyses of short sequences suggest that SARS coronaviruses may have originated in China, however their exact geographical origin is unclear. In order to further investigate the origin of SARS coronavirus and the evolution of SARS-like coronavirus in bats, this study conducted large-scale detection and analysis of SARS-like coronavirus in bat samples collected from twenty regions in China from 2011 to 2016, which is important to reveal the origin of SARS coronavirus and the geographical evolution pattern of SARS-like coronavirus in bats. It is important for revealing the origin of SARS coronavirus and the geographic evolution of SARS-like coronavirus in bats, and provides a basis for the prevention and prediction of emerging coronavirus diseases of bat origin similar to SARS.

Chapter 2 Materials and Methods

2.1 Experimental materials

2.1.1 Materials and reagents

2.1.1.1 Reagents

(1) VTM buffer: Hanks buffer for buffer preparation was obtained from Gino Biologicals; bovine serum albumin (BSA), amphotericin, penicillin and streptomycin were purchased from Sigma.

(2) SuperScript III One-Step RT-PCR with platinum Taq Enzyme Mix, Platinum Taq DNApolymerase: purchased from Invitrogen.

(3) Vector: pGEM-T Easy vector was purchased from Promega.

(4) T4 DNA ligase: purchased from New England Biolabs, Inc.

(5) dNTP mix: purchased from Shanghai Bioengineering Bioengineering Co.

(6) E. coli DH5a receptor cells: this laboratory keeps strains: the laboratory uses receptor E. coli are prepared by themselves.

(7) 2x Taq Master Mix-purchased from Jayosun.

(8) RNAlatter; purchased from Invesco Jetty Trading Co.

(9) Primers: Synthesized by Shanghai Bioengineering Co. or Beijing Kengke Xinye Biotechnology Co.

(10) General reagents such as agarose and sodium chloride are all GMP products.

2.1.1.2 Reagent kit

(1) High pure viral RNA kit: used to extract viral RNA, purchased from Roche.

(2) ENZA⑩Gel Extraction Kit: Used to recover DNA fragments from nucleic acid gels, purchased from OMEGA.

2.1.1.3 Culture medium

(1) LB liquid medium: add 5 g yeast extract, 10 g tryptone, 10 g sodium chloride in double-distilled water in turn, stir well, then fix the volume to l L with double-distilled water and autoclave at 121℃ for 20 min.

(2) LB solid medium: On the basis of prepared LB liquid medium, add 1.5 g agar powder per 100 ml, dispense into 250 ml conical flasks, and autoclave at 121°C for 20 min.

(3) SOC medium: for E. coli recovery culture after electroconversion, 10 g tryptone, 2.5 g yeast extract, 0.25 g NaCl, 0.093 g KCI were added to double-distilled water, stirred with a magnetic stirrer to dissolve completely and then fixed to 490 ml, stirred well and autoclaved at 121°C for 20 min. 10 ml of 1 mol/L glucose solution (5.4 g of glucose was weighed and dissolved in 30 ml of double-distilled water, and then completely dissolved and filtered through 0.22 L filter membrane) and 5 ml of 1 M MgCl solution. After mixing well, dispense into 2 ml EP tubes and store at -20°C. Take out and thaw on a shaker before electrotransformation.

2.1.1.4 Analysis Software

(1) DNAstar Lasergene (v7.1): for sequencing results processing.

(2) MAFFT (v7.304b): for sequence comparison [112].

(3) MEGA7: used for sequence alignment, conversion of nucleotides to amino acids, etc. [113].

(4) WebLogo 3: http: //weblogo.threeplusone.com/create.cgto search the sequence for conserved region and for the design of concise primers.

(5) Primerpremire 5.0: for specific primer design.

(6) DAMBE (v5.38): for saturation detection of sequences [114].

(7) SplitsTree (v4.14.8), RDP4 (Recombination Detection Program v4.8.0): for recombination detection [115,116].

(8) ESpript (v3.0): http: //espript. ibcp. fr/ESPript/cgi-bin/ESPript. cgi, online beautification of multiple comparison sequences [117].

(9) jModelTest (v2.1.7): for nucleotide substitution model testing [118].

(10) PhyML (v3.1): for the construction of phylogenetic trees [119].

(11) TrimGalore (v0.4.5): filter and de-join the reads obtained from high-throughput sequencing.

(12) CLC Genomic Workbench (v11.0.1): for the processing and analysis of high-throughput sequencing data[1]

(13) BLAST (v2.6.0): construct database, perform sequence similarity comparison, etc.

(14) FigTree (v1.4.2), iToL (https://it01.embl.de/): for visualization of phylogenetic trees.

(15) SnapGene (v1.1.3), ORFfinder (https://www.ncbi.nlm.nih.gov/orffmder/): prediction of open reading frames (ORFs).

(16) Jane4, TreeMap3.0b: analysis of cophylogeny between hosts and pathogens [120,121].

(17) BioEdit (v7.25): Edit sequences, convert sequence formats, etc.

2.1.2 Sample collection

Between 2011 and 2016, bat pharyngeal swabs or guano samples were collected from bats in their natural habitats in twenty regions of China, including Yunnan, Guangxi, Guizhou, Guangdong, Chongqing, Hubei, Hunan, Henan, Shaanxi, Shanxi, Jiangxi, and Hainan. Captured bats were released to their roosts after sample collection to minimize harm to them. The collected samples were first placed in a bath containing 1 ml of VTM buffer (Hanks buffer, 1% BSA, 15 μg/ml amphotericin, 100 U/ml penicillin, and 50 μg/ml streptomycin, pH 7.4) or 1 ml RNAlatter in lyophilized tubes, and then stored in liquid nitrogen, transported to the laboratory and stored in a -80°C refrigerator for backup. All sample collection activities were in accordance with the relevant regulations of the Law of the People's Republic of China on the Protection of Wild Animals and the Implementing Regulations of the People's Republic of China on the Protection of Terrestrial Wild Animals, and were approved by the Animal Ethics Committee of Wuhan Institute of Virology, Chinese Academy of Sciences (approval number: WIVH5210201).

2.2 Virus nucleic acid extraction

Viral nucleic acids were extracted in a BSL-2 biosafety cabinet using Roche's High Pure Viral RNA Kit Nucleic Acid Extraction Kit (Roche). Refer to the kit instructions as follows.

(1) Add 400 l Binding Buffer was added to 4 μl Poly(A) Carrier RNA in 1.5 ml centrifuge tubes, mix gently and centrifuge briefly at 4°C for use (within 1 h).

(2) Add 200 μl sample treatment supernatant, vortex shake for 15 s to mix well and leave at room temperature (15-25°C) for 10 min.

(3) Centrifuge briefly, add the above solution (600 μl) into an adsorption column (High Pure Filter Tube), cover with a cap, and centrifuge at 8,000 g for 15 s. Discard the filtrate and collection tube and replace with a new 2 ml collection tube.

\(4) Carefully open the adsorption column, add 500 μl Inhibitor Removal Buffer, cap, centrifuge at 8,000 g for 1 min, discard the filtrate and collection tube, and replace with a new 2 ml collection tube.

(5) Carefully open the adsorption column, add 450 μl Wash Buffer, cap, centrifuge at 8,000 g for 1 min, discard the filtrate and collection tube, and replace with a new 2 ml collection tube.

(6) Carefully open the adsorption column, add 450 μl Wash Buffer, cover, centrifuge at 8,000 g for 1 min, and discard the filtrate and collection tube (the collection tube does not need to be discarded in this step).

(7) Centrifugation at maximum speed (≥13,000 g) for 15 s.

(8) Place the adsorption column into a new 1.5 ml centrifuge tube, add 50 μl Elution Buffer, leave at room temperature for 1 min, centrifuge at 8,000 g for 1 min to collect the viral RNA for direct cDNA synthesis or 25 μl tube for viral RNA splitting, store the RNA at -80°C.

2.3 Coronavirus detection and bat species identification

2.3.1 One-step semi-nested RT-PCR for coronavirus detection

To improve the sensitivity of coronavirus detection, this study used a one-step semi-nested RT-PCR to amplify a segment of approximately 440 bp of the coronavirus RNA-dependent RNA polymerase (RdRp) gene, which is commonly used for coronavirus detection because it is relatively conserved in all four genera of coronaviruses [122]. The primer sequences are shown in Table 2.1 and were synthesized by Shanghai Bioengineering Co.

Table 2.1 Common primers for bat coronavirus detection [122].

表 2.1 蝙蝠冠状病毒检测的通用引物[122]。

Table 2.1 The universal primers for coronavirus detection [122].

引物名称	序列（5'-3'）
CoV-FWD3	GGTTGGGAYTAYCCHAARTGTGA
CoV-FWD4	GAYTAYCCHAARTGTGAYAGAGC
CoV-RVS3	CCATCATCASWYRAATCATCATA

Two rounds of reaction system (both 25 μl) and PCR reaction conditions were as follows: in the first reaction system, 2x Run Mix 12.5 μl, Nuclease-Free Water (NFW) 5.5 μl, MgSO4 0.5 μl, RNase Inhibitor 1.0 μl, primers CoV-FWD3, CoV-RVS3 1.0 μl each l, SuperScript III/Platinum Taq Enzyme Mix 1.0 μl, extracted RNA 3.0 μl. The reaction conditions were: reverse transcription at 50°C for 30 min, pre-denaturation at 94°C for 2 min, denaturation at 94°C for 15 s, annealing at 50°C for 30 s, extension at 68°C for 40 s, and extension at 68°C for 5 min after 40 cycles of reaction. after the first round of reaction was completed, the product obtained in the first round was used as the template for the second round of reaction, and the reaction system was sequentially added: ddH20 17.9 μl, 10x PCR Buffer 2.5 μl, Mgcl2 1.0 μl, dNTP Mix 0.5 μl, primers CoV-FWD4 and CoV-RVS3 1.0 μl each, Platinum Taq DNA polymerase 0.1 μl, 1 μl of the first round product. Reaction conditions: after pre-denaturation at 94°C for 2 min, denaturation at 94°C for 20 s, annealing at 50°C for

30 s, extension at 72°C for 30 s, reaction for 40 cycles followed by extension at 72°C for 5 min.

The obtained PCR products were run on 2% agarose gel electrophoresis, and PCR amplification products with a single target band size were sent directly to Kuntai Rui Biotechnology Co. For samples with weak or non-specific amplified bands and samples with overlapping peaks in the peak map obtained by direct PCR sequencing, the PCR products were purified using Gel Extraction Kit (Omega), and the purified products were ligated onto pGEM-T Easy vector after purification. The ligation system was prepared according to the T4 ligase instructions, and the ligase was added to the 10 μl of ligation system, add T4 Ligation Buffer 0.5 μl, pGEM-T Easy vector 0.3 μl, T4 DNA Ligase 0.5 μl, purified product 3.7 μl, and ligated in a water bath at 16°C overnight. Prior to electrotransformation, the SOC liquid medium placed in the -20°C refrigerator was taken out and thawed, and then 1.5 μl ligation product was added to 60-80 μl DH5 receptor cells on ice, mix well, add the mixture to the electrotransformation cup, cover the lid, try not to inhale air bubbles, dry the surrounding of the transformation cup with absorbent paper, and then gently put it in the electrotransformer for electrotransformation (voltage 1800 V), then add the warmed SOC liquid medium to the electrotransformation cup on the ultra-clean table, resuspend the cells, and transfer all the liquid in the electrotransformation cup to the original tube where the SOC liquid medium was put. After recovery for one hour at 37°C and 220 rpm on a shaker, 200 gl of transformation product was applied to ampicillin-resistant medium and incubated overnight at 37°C in an inverted oven. A dozen white colonies on each transformation plate were selected for colony PCR identification using universal primers M13-F (5'-TGTAAAACGACGGCCAGT-3') and M13-R (5'-CAGGAAACAGCTATGACC a 3 ') were tested in a 15 μl of colony PCR system was added sequentially: 2x Taq master Mix 7.5 μl, ddH20 5.5 μl, M13-F and M13-R 1.0 μl each. The PCR reaction system was: pre-denaturation at 95°C for 5 min, denaturation at 95°C for 30 s, annealing at 50°C for 30 s, extension for 1 min, 35 cycles of reaction and extension at 72°C for 7 min. clones with the target band size were selected and sent for sequencing, no less than 5 clones per sample. The sequencing results of the company were processed with SeqMan of DNAstar Lasergene (v7.1) with reference to the sequencing peak maps, and the sequences obtained were directly blastn online in the Blast interface of NCBI (https: //blast.ncbi.nlm.nih.gov/Blast.cgi) to identify the virus species. The detected sequences were considered to be SARS-like coronaviruses if the similarity between them and the existing SARS-like coronaviruses in GenBank was greater than or equal to 95%.

2.3.2 Bat species identification

Bat species were initially identified based on bat morphology and subsequent molecular species identification was performed by amplifying the cytochrome b (CytB) gene or the reduced nicotinamide adenine dinucleotide dehydrogenase subunit 1 (ND1) gene in bat mitochondria, with experimental steps referring to the methods published by Irwin et al. and Mayer et al. [123,124].

2.4 Amplification of full-length RdRp, S and ORF8 genes

All bat SARS-like coronavirus sequences with full-length genomes in GenBank were downloaded and then aligned with MAFFT (v7.3.04b) to intercept the full-length RdRp, S and ORF8 genes in MEGA7, respectively, and then WebLogo 3 was used to find the conserved regions of the sequences and design the simplex primers, which amplified fragments excluding the coronavirus The 440bp fragment detected [112, 113]. Because the plausible read length of one reaction for sanger sequencing does not exceed 800 bp, to ensure the accuracy of sequencing the amplified sequences, we amplified both the full-length RdRp and S in three segments. The simplified primers are shown in Table 2.2. The unamplified fragments were amplified with Primer premire 5.0 by redesigning the specific primers based on the already amplified fragments as shown in 2.3.1. If the primers are not indicated for the first and second rounds, the same primers were used for both rounds of PCR.

Table 2.2 Degenerate primers for SARS-like coronavirus RdRp, S and ORF8 genes.
Table 2.2 Degenerate primers of RdRp, S and ORF8 genes for SARSr-CoV

Primer name	Primer sequence (5'-3')		Annealing temperature	Amplification length
7896-S-1-1	S-1-F1	TTCTCATGGTGGACWGCBTTTG	50°C	1526bp
(1st round)	S-1-R	TARTCWGCAATAACACCRGT		
7896-S-1-2	S-1-F2	ATTGAYGGHTATACCATGCAYG	48°C	1431bp
(2nd round)	S-1-R	TARTCWGCAATAACACCRGT		
7896-S-2	S-2-F	ABTGTGTTGCDGAYTAYWCT	48°C	1582bp
	S-2-R	ACCTATADGCCATYTGCATAGC		
7896-S-3	S-3-F	GCWAGAGAYCTYATYTGTGC	48°C	1350bp
	S-3-R	GTAGCATGAACAGTACTTGC		
8363-R-1-1	R-1-F1	AACCTAAATAGAGGTATGGT	42°C	1500bp
(1st round)	R-1-R	GTTTACATCCTGATTATGTAC		
8363-R-1-2	R-1-F2	CAAGTGGWGGACAACCAATCAC	44°C	1402bp
(2nd round)	R-1-R	GTTTACATCCTGATTATGTAC		
8363-R-2-1	R-2-F	CATTGTGCAAACTTTAATGT	45°C	1154bp
(1st round)	R-2-R1	ACACTATTAGCATAAGCAGTTGT		
8363-R-2-2	R-2-F	CATTGTGCAAACTTTAATGT	45°C	1121bp

(2nd round)	R-2-R2	GATGATGTTCCACCTGGTTT		
8363-R-3	R-3-F	TTAAGTGAGATGGTCATGTG	46°C	962bp
	R-3-R	TTGCAAACATAGGGATTAACAG		
ORF8-1-1	M-8-F1	TAGGAGCKTCGCAGCGTGTAG	59°C	1513bp
(1st round)	N-8-R1	CCACGAACTCGTCGGGTAGCT		
ORF8-1-2	M-8-F2	CAATATTGCTTTGCTAGTACAGT	49°C	1302bp
(2nd round)	N-8-R2	CCATGCTGAGTGAGAGCTGTG		

2.5 Whole genome sequencing of bat SARS-like coronavirus

The viral whole-genome sequencing of four bat SARS-like coronaviruses was performed using the Illumina high-throughput sequencing platform in combination with the viral content of the bat SARS-like coronavirus positive samples, the sampling locations and whether the RBD was missing, and the sequencing was performed using the Illumina Truseq Stranded mRNA library building method with the following brief steps.

(1) mRNA fragmentation. The fragmentation system was: 5 μl RNA plus 14 μl fragment primer finish mix, gently pipette 6 times to mix well, then react at 94℃ for 4 min, stabilize to 4℃ and remove, then continue the synthesis of the first strand cDNA.

(2) First-strand cDNA synthesis. Formulate the first strand cDNA synthesis system, the system is: 8 μl First StrandSynthesis Act D Mix+1 μl SuperSript II, mix well with a pipette gun; add 8 μl configured mix, mix well by pipette. PCR reaction program was 25℃/10 min, 42℃/15 min, 70℃/15 min, 4℃/+∞, where the hot cap temperature was set to 100℃.

(3) Second strand cDNA synthesis. Preparation of Second Strand cNDA Synthesis Mix: 20 μl Second Strand Making Master Mix, 5 μl Resuspension Buffer; add the mix to the product of step 2 and mix thoroughly and uniformly with a pipette.

(4) Add "A" tail at the 3' end. The A-tailing system is: DNA 16 μl plus A-tailing Mix 9 μl, total volume 25 μl. The PCR reaction procedure was 37℃/30 min, 70℃/5 min, 4℃/+∞, and then placed on ice after stabilization to prevent inverse reaction.

(5) Splice ligation. The ligating system is: DNA 25 μl, Resuspension buffer (RSB) 2.5 μl, Ligation Mix 2.5 μl and Adapter Index 2.5 μl, total volume of 32.5 μl. After the PCR instrument was finished, 3.5 μl of Stop ligation buffer was added respectively. Mix well by pipetting up and down.

(6) Add 36 μl of well-mixed beads, pipette up and down 10 times; after a warm bath at room temperature for 8-10 min, let stand on a magnetic rack at room temperature for 5 min to remove all supernatants; add 200 μl 80% anhydrous ethanol and let stand at room temperature for 30 s to remove all supernatants and ensure no residue; repeat the ethanol washing step to remove supernatants and then add 21 μl RSB, mix well by blowing, and let

stand at room temperature for 4 min, then let stand at room temperature for 5 min on a magnetic stand; transfer 20 μl to the new wells.

(7) PCR enrichment. The system was: DNA 20 μl, PCR primer Cocktail 5 μL, master Mix 25 μl, total volume of 50 μl; the reaction program was: 98°C/30 s, 98°C/10 s, 60°C/30 s, 72°C/30 s, 10 cycles, 72°C/300 s, 10°C/+∞.

(8) PCR enrichment. The system was: DNA 20 μl, PCR primer Cocktail 5 μL, master Mix 25 μl, total volume of 50 μl; the reaction program was: 98°C/30 s, 98°C/10 s, 60°C/30 s, 72°C/30 s, 10 cycles, 72°C/300 s, 10°C/+∞.

(9) The constructed libraries were qualified by identification (nucleic acid gel electrophoresis or fluorescence quantitative PCR) and sequenced by Illumina sequencer.

Raw data data obtained from high-throughput sequencing were first filtered out with TrimGalore (v0.4.5) to get clean data with low-quality reads and contigs, and then CLC Genomic Workbench (v11.0.1) was used to sequentially perform quality control (fastqc), filtering (trim reads) and de novo assembly, and the obtained contigs were viewed in the graphical interface of CLC Genomic Workbench for the specific assembly. Meanwhile, we used makeblastdb in blast (v2.6.0) software to build a library for all coronaviruses in GenBank with the name of CoV_db, and the contigs obtained earlier were compared to CoV_db as query to blastn, and then the sequences of the contigs on the comparison were extracted. If the length of the longest contig is greater than 20kb, this contig will be directly online blastn, and the sequence with the highest similarity to it is selected as the reference sequence. If the length of the longest contig is less than 20kb, the first 10 long contigs are extracted and blastn online, and the sequences with high similarity to them are found as reference sequences. Then the join contigs module of CLC Genomic Workbench was used to join each sample contigs according to the reference sequence (the software allows adding multiple reference sequences), and the positions that could not match (mapping) to the reference sequence existed gaps, and these gaps were denoted by N. Later, according to the sequence of both ends of the gaps PCR amplification and sanger sequencing were carried out by designing specific primers based on the sequence of the two ends of the gaps, so as to complete the gaps and obtain the full-length genome of the virus.. The open reading frames of the viral genome were predicted using the online ORF finder (https: //www. ncbi. nlm. nih. gov/orffinder/) and SnapGene (v1.1.3).

2.6 Sequence processing and analysis

The nucleotide sequences of bat SARS-like coronaviruses obtained in this study were aligned first with MAFFT (v7.304b) with the nucleotide sequences of SARS and SARS-like coronaviruses downloaded from Genbank, then the sequences were manually processed in MEGA7, converted to sequence file format in BioEdit (v7.25), and finally ESpript (v3.0) to beautify and display the multiple comparison sequences [112,113,117].

2.7 Phylogenetic analysis of SARS-like coronaviruses

The nucleotide sequences after 2.6 processing were subjected to phylogenetic analysis. The saturation of the sequences was first detected by Xia'test in DAMBE (v5.38), and if Iss<Issc and the difference was significant (P<0.05), it indicated that the nucleotide sequence substitutions were not saturated and could be used to construct a phylogenetic tree [114]. Then, the nucleotide substitution model was tested with iModelTest (v2.1.7), and the result of AIC (Akaike Information Criterion) was used as the standard for the best nucleotide substitution model, and the phylogenetic tree was constructed by the maximum likelihood (ML) method in PhyML (v3.1) with the nucleotide substitution model set up, bootstrap set to 1000, and other parameters defaulted [118,119]. The constructed evolutionary trees were visualized in FigTree (v1.4.2) and iTOL (https: //it01. embl. de/).

2.8 Recombination Detection and Analysis

According to previous reports in the literature [87], the S gene is a significant recombination hotspot for SARS-like coronaviruses. The full-length S gene was analyzed by Phi test and the presence of recombination was determined by the P value, and if this result indicated the presence of significant recombination, further validation was performed with RDP4 (v4.8.0) [115,116]. Suspected recombinants were identified simultaneously with RDP, GENECONV, BootScan, Maxchi, Chimaera, SiSscan, and 3Seq programs using Bonferroni correction with the significance level set at 0.05. Only when four or more of the seven detection programs showed significant recombination (P<1.0×10 '-6), the virus was considered to be a recombinant strain. To visualize the S recombination, the aligned S1 gene sequence was first manually divided into two fragments S1-NTD and S1-RBD according to the recombination site, and then the evolutionary tree was constructed for these two fragments separately by ML method, and finally the same nodes were joined to form a tanglegram.

2.9 Coevolutionary analysis between SARS-like coronaviruses and their hosts

To investigate the co-evolution of bat SARS-like coronavirus and its host, we performed co-phylogeny analysis in Jane4 based on events, with the phylogenetic tree of SARS-like coronavirus built on the full-length RdRp nucleotide sequence and the phylogenetic tree of the host built on the full-length CytB gene nucleotide sequence of the host, while setting the "cost" of each event in Jane4: co-divergence=0, duplication=1, host switch=1, loss=1, failure to diverge=1. The generation and population size were set to 100, and the estimated "cost" was determined by matching the estimated "cost" was compared with the zero (null) distribution obtained from 100 random mapping calculations to obtain the significant type of coevolution [120]. The coevolutionary relationship between SARS-like coronaviruses and their hosts can be further verified and demonstrated by TreeMap3.0b [116].

Chapter 3 Results and Analysis

3.1 Prevalence of SARS-like coronavirus in bats

During 2011-2016, a total of 9261 bat anal swabs or fecal samples were collected from twenty regions in China (Yunnan, Guangxi, Guizhou, Guangdong, Hainan, Chongqing, Sichuan, Hunan, Hubei, Jiangxi, Henan, Jiangsu, Zhejiang, Shandong, Hebei, Beijing, Shanxi, Shaanxi, Gansu, and Ningxia), as shown in Table 3.1. A total of 170 SARS-like coronaviruses were detected from these bat samples by coronavirus testing, with a positive rate of 1.84%. Molecular species identification of these bat SARS-like coronavirus-positive samples showed that the positive samples were mainly from bats of the family Rhinolophidae, which including R. affinis (R. affinis), and R. ferrumequinum (R. ferrumequinum), and R. macrotis (R. macrotis), and R. pearsonii (R. pearsonii), and R. pusillus (R. pusillus), and Chinese chrysanthemum-headed bat (R. sinicus) and the remaining five positive samples were from the genus Rhinolophus, which could not be identified to species. The remaining five positive samples were from A. stoliczkanus and R. pusillus of the Hipposideridae family. stoliczkanus and H. pratti. In addition, from a regional perspective, positive samples were found mainly in nine regions, namely Yunnan, Guangxi, Guizhou, Guangdong, Chongqing, Hunan, Hubei, Shanxi and Beijing, among which, SARS-like coronavirus positive samples were detected for the first time in three regions, namely Chongqing, Hunan and Beijing, while SARS-like coronavirus was not detected in the other eleven regions.

3.2 Phylogenetic analysis of SARS-like coronaviruses

Regarding the phylogenetic study of SARS-like coronavirus, most previous studies were based on a conserved 440-bp RdRp sequence used for detection, which may not be accurate for inferring the evolutionary history of SARS-like coronavirus due to the short sequence and limited genetic information covered. Therefore, in this study, the full-length RdRp sequence (2757 bp) was amplified for phylogenetic analysis of the virus, and finally 60 full-length RdRp sequences of SARS-like coronaviruses were obtained from positive samples. The amplified full-length RdRp sequences were compared with the full-length RdRp sequences of SARS coronavirus and SARS-like coronavirus in GenBank, and the phylogenetic tree was constructed by the maximum likelihood method, using human MERS coronavirus HCoV-EMC (JX869059) as the outgroup of the phylogenetic tree.

Table 3.1 Bat SARS-like coronaviruses in 20 regions of China from 2011 to 2016 XXXX

Table 3.1 Prevalence of SARSr-CoV in twenty regions in China during 2011-2016

Sampling location	Sample type	Number of Samples	Host species (Latin name)

		Total number of samples	SARSr-CoV Positive	
Yunnan	Anal swab, fecal pellet	2815	92	R. sinicus, R. affinis, R. ferrumequinum, R. spp, Aselliscus stoliczkanus
Guangxi	Anal swab	612	11	R. sinicus, R. pusillus
Guizhou	Anal swab	440	16	R. sinicus
Guangdong	Anal swab	1979	33	R. sinicus, R. pusillus, R. affinis, H. pratti
Hainan	Upper anal swab	458	0	
Chongqing	Anal swab, fecal pellet	191	1	R. macrotis
Sichuan	Anal swab, fecal pellet	396	0	
Hunan	Anal swab	208	5	R. sinicus
Hubei	Anal swab, fecal pellet	1383	3	R. pearsonii
Jiangxi	Anal swab	48	0	
Henan	Anal swab, fecal pellet	335	0	
Jiangsu	Anal swab	44	0	
Zhejiang	Anal swab	48	0	
Shandong	Anal swab	49	0	
Hebei	Anal swab	48	0	
Beijing	Anal swab	30	1	R. ferrumequinum
Shanxi	Anal Swab	122	8	R. ferrumequinum
Shaanxi	Anal Swab	25	0	
Gansu	Anal Swab	13	0	
Ningxia	Anal Swab	17	0	


Lineage 1
Lineage 4
Lineage 3
Lineage 2

Figure 3.1 Phylogenetic analysis of SARS-like coronavirus and SARS coronavirus based on the complete RdRp sequence.

The phylogenetic tree was constructed using the ML method based on the complete RdRp nucleotide sequence and visualized in iTOL (https://itol.embl.de/). The human MERS coronavirus HCoV-EMC (JX869059) was used as the outgroup in the evolutionary tree and is marked in gray in the evolutionary tree. SARS-like coronaviruses found in this study are marked in bold. branches of SARS coronaviruses are marked in red. The confidence level of each node on the evolutionary tree is indicated by the size of the purple circle to indicate its value. The geographical distribution of SARS-like coronaviruses in lineage 1, lineage 2, lineage 3 and lineage 4 is indicated by different colored dots at corresponding locations on the map. The regional abbreviations are as follows: YN: Yunnan; GZ, Guizhou; GX, Guangxi; GD, Guangdong; HK, Hong Kong; HuN, Hunan; CQ, Chongqing; HuB, Hubei; HeN, Henan; SAX, Shaanxi; SX, Shaanxi; HeB, Hebei; BJ, Beijing; zJ, Zhejiang; er, Jilin.

Fig 3.1 Phylogenetic analysis of SARSr-CoVs and SARS-CoVs based on the complete RdRp sequences.

The phylogenetic tree was made using the Maximum-likelihood approach based on the complete RdRp sequences and viewed in iTOL(https://itol.embl.de/). The human MERS-CoV HCoV-EMC (JX869059) was used as an outgroup and shown in gray in that tree. The SARSr-CoV strains discovered in this study are shown in bold. The branch of SARS-CoV is marked in red. The bootstrap values of the node are displayed with the size of the purple circle. The geographic distribution of SARSr-CoVs from lineage 1, lineage 2, lineage 3 and lineage 4 is indicated by colored dots at the corresponding locations on the map. Location abbreviations are as follows YN, Yunnan; GZ, Guizhou; GX, Guangxi; GD, Guangdong; HK,

Hong Kong; HuN, Hunan; CQ, Chongqing; HuB, Hubei; HeN, Henan; SAX, Shaanxi; SX, Shanxi; HeB, Hebei; BJ,Beijing; ZJ, Zhejiang; JL, Jilin.

The evolutionary tree of RdRp (Figure 2.1) shows that SARS-like coronaviruses tend to cluster together according to sampling locations rather than according to host species, suggesting that geographical location plays an important role in the evolution of SARS-like coronaviruses and that there may be mutual exchange between SARS-like coronaviruses from neighboring regions. As a whole, the evolutionary tree of RdRp consists of four lineages, lineage 1, also known as the southern lineage, consists mainly of SARS-like coronaviruses from Yunnan, Guizhou, Guangxi, Guangdong and Chongqing, as well as human and civet SARS The SARS-like coronaviruses in this lineage are closer to human SARS coronaviruses in the evolutionary tree than SARS-like coronaviruses in other lineages, in which the SARS-like coronaviruses with the closest affinity to SARS coronaviruses are from Yunnan, followed by Guangxi, Guizhou, Guangdong and Chongqing; SARS-like coronaviruses from Guizhou, Guangdong, Hunan, Hubei, Zhejiang and Hong Kong form the south lineage 2. SARS-like coronaviruses from Hubei, Shaanxi, Shanxi, Henan, Hebei, Beijing, and Jilin comprised lineage 3; and lineage 4 was composed of SARS-like coronaviruses from Yunnan. In addition, a SARS-like coronavirus strain BM48-31 (GU190215) from Bulgaria, Europe, was between lineage I and lineage 4 in the evolutionary tree. Yunnan has both SARS-like coronaviruses that are closely related to human SARS coronaviruses (Lineage 1) and SARS-like coronaviruses that are distantly related to human SARS coronaviruses (Lineage 4), indicating that the genetic diversity of SARS-like coronaviruses in this part of Yunnan is relatively high. Although human SARS coronavirus was first outbreak in Guangdong, the closest to human SARS coronavirus in the southern lineage is not from Guangdong, indicating that Guangdong is not the ancestral origin of SARS-like coronavirus. Combined with the map below the evolutionary tree, SARS-like coronaviruses from Yunnan, Guizhou and Guangdong were distributed in two or more lineages, indicating that SARS-like coronavirus diversity was higher in these three regions than in others.

3.3 Genomic analysis of SARS-like coronaviruses

Based on the RdRp evolutionary tree, four SARS-like coronaviruses (Ra4991_Yunnan, Rs5725_Yunnan, Rs8561_Guangdong, Rs141456_Guangxi) were selected for viral whole-genome sequencing, taking into account the viral content of the samples, the sampling locations and whether the RBD was missing. The first strain was from the Chinese chrysanthemum-headed bat, and the other three were from the Chinese chrysanthemum-headed bat. The full-length viral genomes were obtained for all three strains except for Rs5725_Yunnan, which had a partial gap in the nucleocapsid protein gene that was not fully complemented and was attempted to be amplified, but still not amplified. The open reading frames (ORF1a, S, ORF3, E, M, ORF6, ORF7, ORF8, N) of these four viruses and other SARS-like coronaviruses were compared with the amino acid sequence of human SARS coronavirus Tor2 (AY274119) (Table 3.2), and it was found that Rs5725_Yunnan and WIV1, also from Yunnan (KF367457) and WIV16 (KT444582), all ORFs (except ORF8) had high amino acid similarity with human SARS coronavirus, and the amino acid similarity of ORF8 of Rs5725_Yunnan (31.1%) was higher than both of them, but not as high as another strain of Yunnan virus Rs4084 (KY417144) (85.9%), however, the amino acid sequence of S gene of this strain was not very similar to that of human SARS coronavirus (89.8%).

Table 3.2 Comparison of ORF amino acid homology between bat SARS-like coronavirus and human SARS coronavirus Tor2.

Table 3.2 Comparison of ORF amino acid identities between bat SARSr-CoVs and human SARS-CoV Tor2.

CoV	ORF1a	S	ORF3	E	M	ORF6	ORF7	ORF8	N
LYRa11/Yunnan	95.8	89.2	90.8	98.6	97.7	95.2	94.2	15.4	97.8
WIV16/Yunnan	98.0	96.8	97.0	100.0	98.1	92.0	95.0	16.1	99.5
WIV1/Yunnan	97.9	92.1	96.7	100.0	98.1	93.6	95.0	16.9	99.7
Rs4084/Yunnan	97.9	89.8	97.0	98.6	98.1	96.8	95.9	85.9	99.5
Anlong-103/Guizhou	98.1	79.4	89.7	100.0	98.6	96.8	95.0	31.8	98.3
Rp3/Guangxi	96.5	78.5	83.5	100.0	97.2	92.0	95.0	15.4	97.8
HKU3-7/Guangdong	94.0	78.1	82.8	98.6	97.2	92.0	93.4	15.4	98.1
HKU3-1/Hong Kong	94.1	77.9	81.7	100.0	98.6	93.6	94.2	14.7	96.4
Rf1/Hubei	94.0	76.6	86.1	96.0	97.7	93.6	91.8	23.9	95.4
Jiyuan-84/Henan	94.0	76.8	86.8	96.0	97.7	92.0	94.2	25.3	96.2
Rp/Shaanxi	95.4	79.0	82.4	97.3	96.8	92.0	92.6	15.4	97.8
HeB2013/Hebei	94.1	76.8	87.2	96.0	97.7	93.6	93.4	25.3	96.2
JL2012/Jilin	93.2	74.6	86.1	97.3	97.2	87.3	90.1	NA	95.7
Ra4991_Yunnan	80.0	75.4	73.8	94.7	90.9	66.6	83.6	12.6	90.5
Rs5725_Yunnan	98.2	94.8	97.0	100.0	100.0	100.0	95.9	31.1	NC
Rs8561_Guangdong	80.4	75.2	70.5	94.7	89.6	68.2	83.6	12.6	92.8
Rs141456_Guangxi	96.1	78.6	83.5	98.6	98.6	88.8	94.2	16.3	98.1
BM48-31/Bulgaria	80.8	74.7	69.0	92.1	88.5	49.2	56.5	NP	87.2

NA, not available; NP, not present; NC, not complete.

The virulent strains obtained in this study were coarsely labeled and their values were marked in red when the amino acid sequences of S, ORF3 and ORF8 of bat SARS-like coronavirus were greater than 90%, 90% and 85% amino acid similarity to the corresponding regions of human SARS coronavirus, respectively.

3.4 Analysis of S-gene diversity and recombination in SARS-like coronaviruses

The S gene of SARS coronaviruses is critical for virus entry into host cells and can determine the host range, especially the receptor binding region (RBD) of the S gene. Compared with the amino acid sequence of the receptor binding motif (RBM) of human SARS coronavirus RBD (Figure 3.2), a few

bat SARS-like coronaviruses do not have deletions in the RBM region, including a virus strain Rs5725_Yunnan obtained in this study, and SARS coronaviruses have 67.9%-96.1‰ similarity in the RBM region while the vast majority of bat SARS-like coronaviruses had deletions in this region, except for the European virus BM48-31, which had only one deletion, while all others had two deletions. SARS-like coronaviruses without deletions in the RBM region are able to utilize ACE2 receptors in bats, civets, and humans, whereas SARS-like coronaviruses with deletions in this region are not able to utilize ACE2 receptors, and therefore these viruses may not be the direct ancestors of human SARS coronaviruses [90]. Currently, bat SARS-like coronaviruses without deletions in the RBM region are only found in Yunnan, where the ancestor of the S gene of human SARS coronaviruses is presumed to have originated.

Recombination is common in coronaviruses, especially in the S gene, and may lead to changes in the host profile of the virus. Recombination of the full-length S gene was detected using the Phi test, which showed the presence of significant recombination. S recombination was then further detected using seven procedures of RDP4, and a total of 31 recombination events were detected (P<1.0×10'-6). The recombination events were found mainly in SARS-like coronaviruses of family lineage 1 and family lineage 2, indicating that frequent recombination occurred in SARS-like coronaviruses of family lineage 1 and family lineage 2.

To visualize the recombination of S genes, a tanglegram was constructed by connecting the same nodes of S1-NTD and S1-RBD evolutionary trees. From Figure 3.3, there are frequent recombination events of S genes, for example, Rs151388_Guizhou is clustered with Rs151262_Guizhou in the NTD evolutionary tree on the left, while they are farther apart in the right RBD evolutionary tree, and they are clustered with Rs151200_Hunan and Rs151161_Hunan, respectively, indicating that recombination may exist between these four viruses, and the recombination event was detected by all seven programs of RDP4 as significant recombination (P<5.08×10'12) .

Figure 3.2 Amino acid sequence comparison results of SARS coronavirus and SARS-like coronavirus receptor binding motif (RBM).

The SARS-like coronavirus found in this study is indicated by the vertical line on its left side, where Rs5725 Yunnan is the sample without deletion in the RBM region, indicated by the red pentagram.

Figure 3.2 Amino acid sequence alignment of the receptor-binding motif (RBM) of SARS-CoVs and SARSr-CoVs.

SARSr-CoVs identified in this study are indicated with the vertical line on the left and the strain Rs5725_Yunnan newly discovered without any deletions in this RBM region is marked with red asterisks

SARSr-CoVs identified in this study are indicated with the vertical line on the left and the strain Rs5725_Yunnan newly discovered without any deletions in this RBM region is marked with red asterisks.

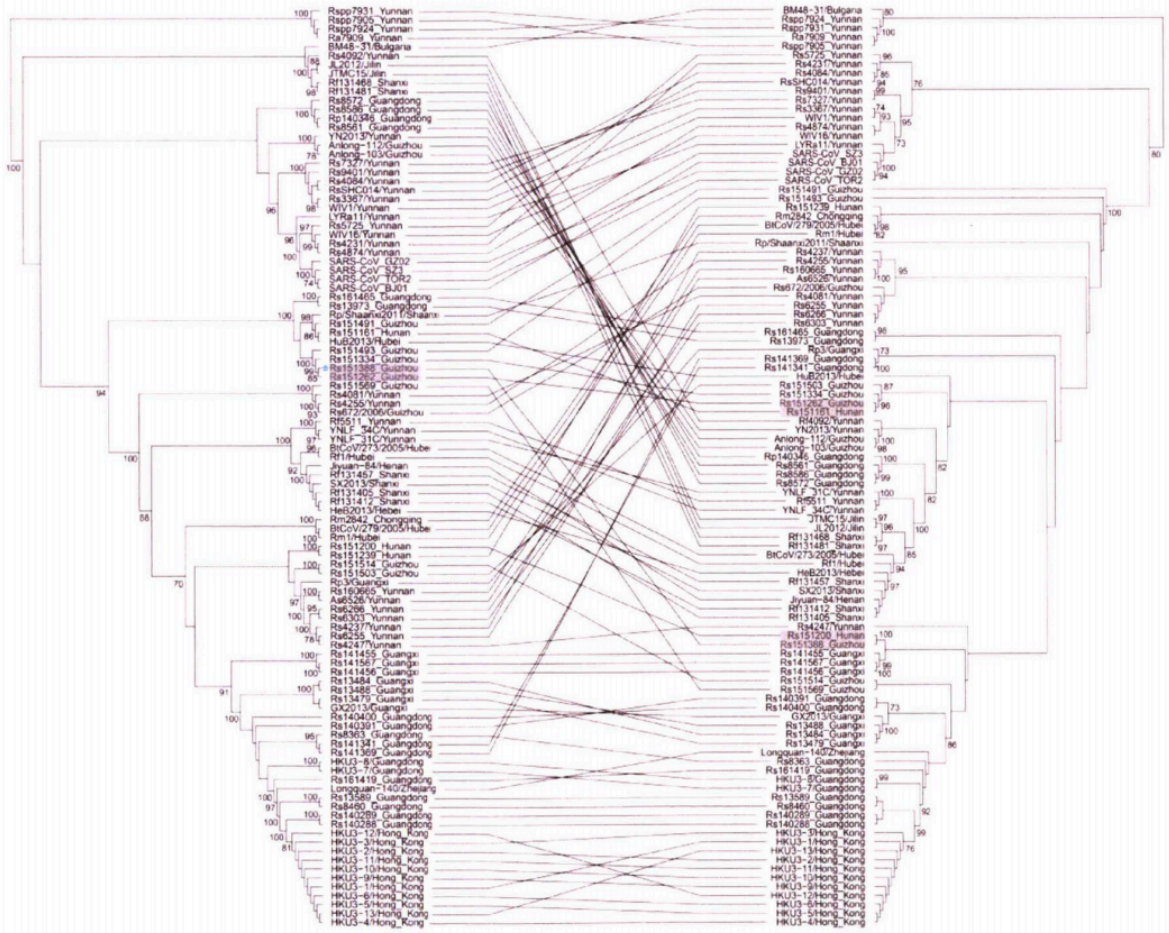


图 3.3 连接 NTD 和 RBD 之间的相同分类单元的进化树。

Figure 3.3 Evolutionary tree connecting the same taxonomic units between NTD and RBD.

On the left is the evolutionary tree constructed by NTD nucleotide sequences, and on the right is the evolutionary tree constructed by RBD nucleotide sequences. Rs151388_Guizhou as an example is marked with a blue pentagram and the associated recombinant virus is shaded in gray.

Figure 3.3 Tanglegram connecting the same taxa between the phylogenetic trees of NTD and RBD region.

The phylogeny on the left was estimated using nucleotide sequences of NTD and the tree on the right was inferred for the RBD sequences. An example of Rs151388_Guizhou is marked with a blue pentagram and relevant recombinant viruses are grey shaded.

3.5 ORF8 gene diversity and phylogenetic analysis of SARS-like coronavirus

The gene deletion phenomenon was also found in the ORF8 gene [87]. As shown in Figure 3.4, we can see that the civet SARS coronavirus and most of the early human SARS coronaviruses are free of deletions in the ORF8 region, mainly in the middle and late stages with 29nt or 82nt deletions, and even some complete deletions of ORF8 in the late stages, due to partial deletions interrupting the encoding of the open reading frame, and ORF8 cleavage into two small open reading frames ORF8a and ORF8b, which may be related to the jump of SARS coronavirus from animals to humans. In contrast, among the SARS-like coronaviruses from bats, except for BM48-31, which has a complete deletion of ORF8, other viruses with partial deletion were not found to have exactly the same deletion as human mid- and late-stage SARS coronaviruses. Rs4084 from Yunnan and Rs151334_Guizhou from Guizhou had 5nt and 4nt deletions, respectively, and ORF8 was divided into ORF8a and 0RF8b. Ra7909_Yunnan from Yunnan and HKU3-8 (GQ153543) from Guangdong had 7nt and 26nt deletions, respectively, and ORF8 was divided into ORF8a, ORF8b and ORF8b. ORF8b and ORF8c.

As can be seen in Figure 3.5, the vast majority of SARS-like coronaviruses have an intact ORF8 without deletions and are a complete ORF8 with a length of 366 nt or 369 nt. Considering that deletions lead to early termination of the reading frame, ORF8b of the human SARS coronavirus Tor2 was chosen as the reference sequence to construct a phylogenetic tree of ORF8, as shown in Figure 3.6. ORF8 evolutionary tree includes two main branches (cluster1 and cluster2) and an independent virus Rs13589_Guangdong. Cluster1 is composed of civet and human SARS coronaviruses and bat SARS-like coronaviruses with complete or partially missing ORF8, and cluster2 is both composed of bat SARS-like coronaviruses with complete ORF8 bat SARS-like coronaviruses. Compared with cluster2, the bat SARS-like coronaviruses in cluster1 were more closely related to human SARS coronaviruses, and these viruses were also found to be from Yunnan, Guizhou, Guangxi and Guangdong, which is consistent with family 1 in the previous RdRp evolutionary tree, further demonstrating that the direct ancestor of human SARS originated from family 1.

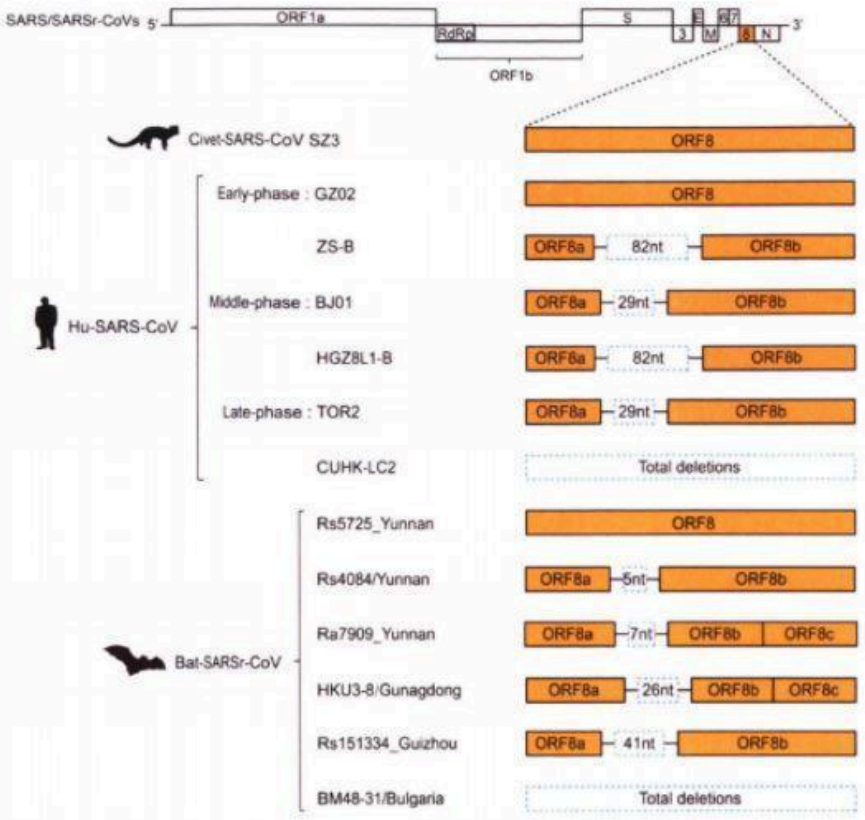
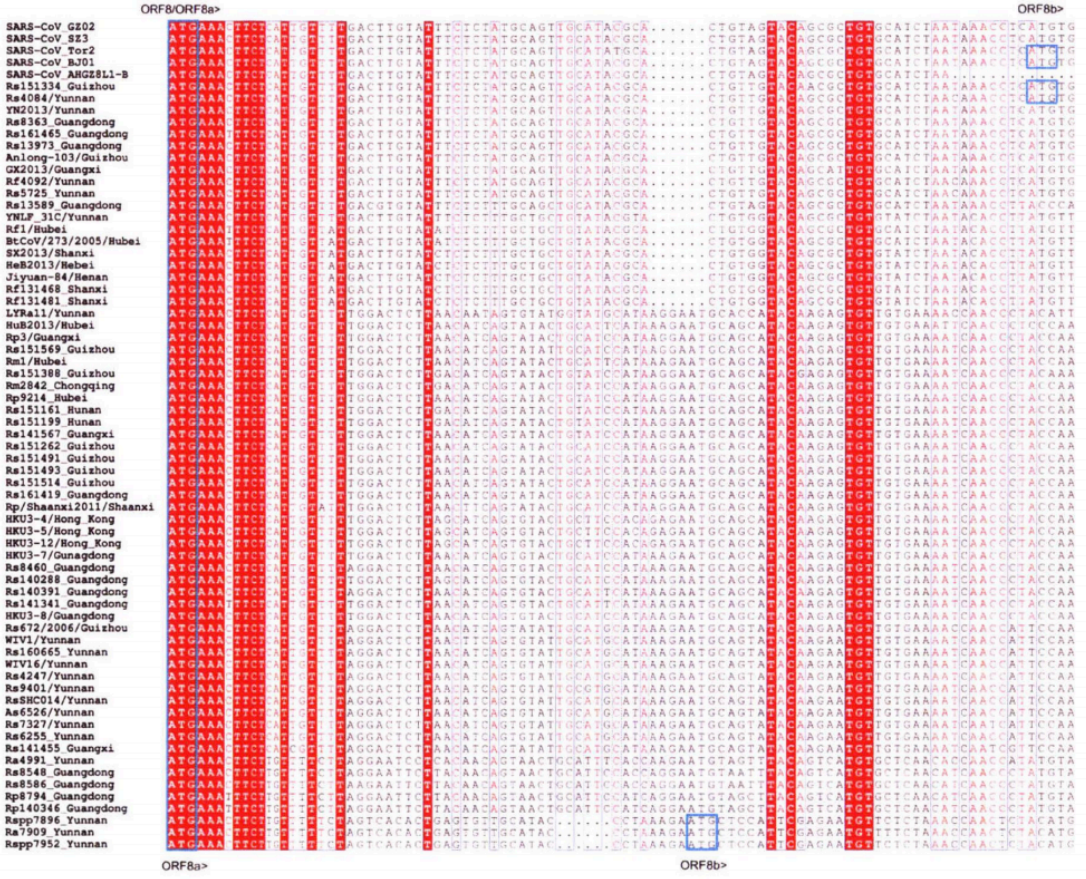
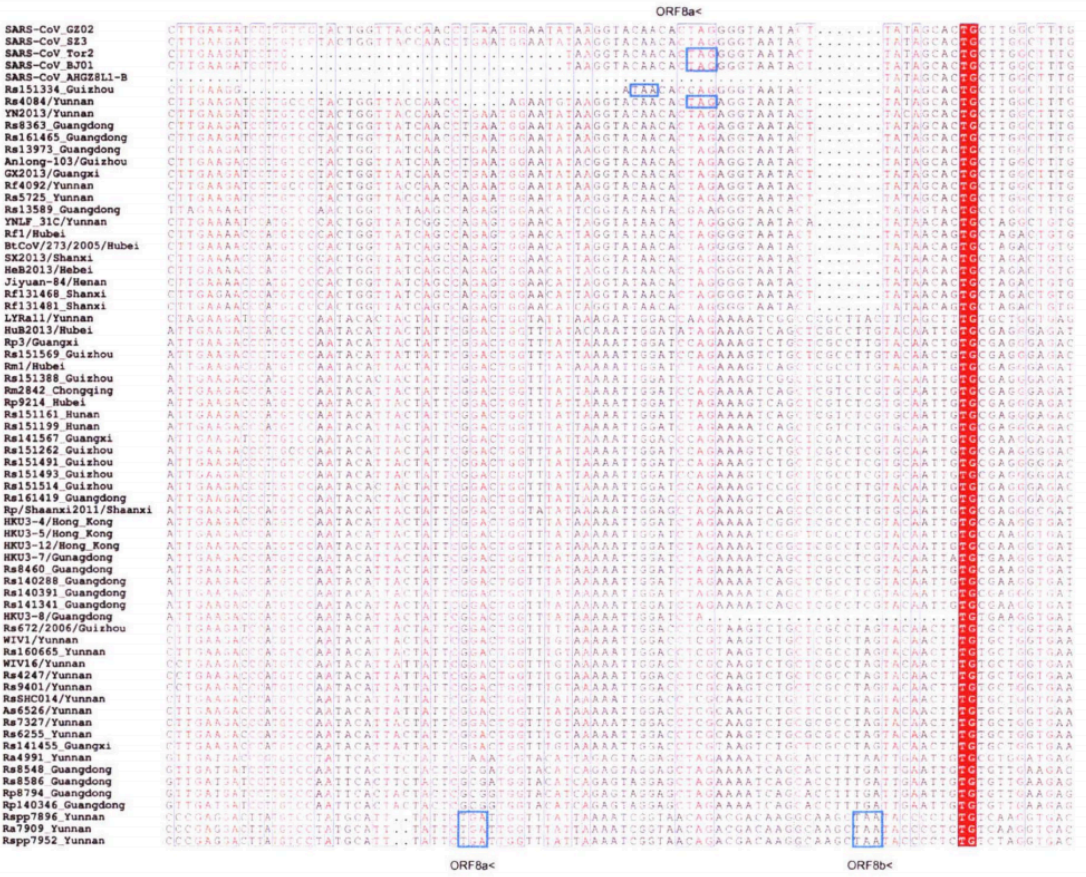


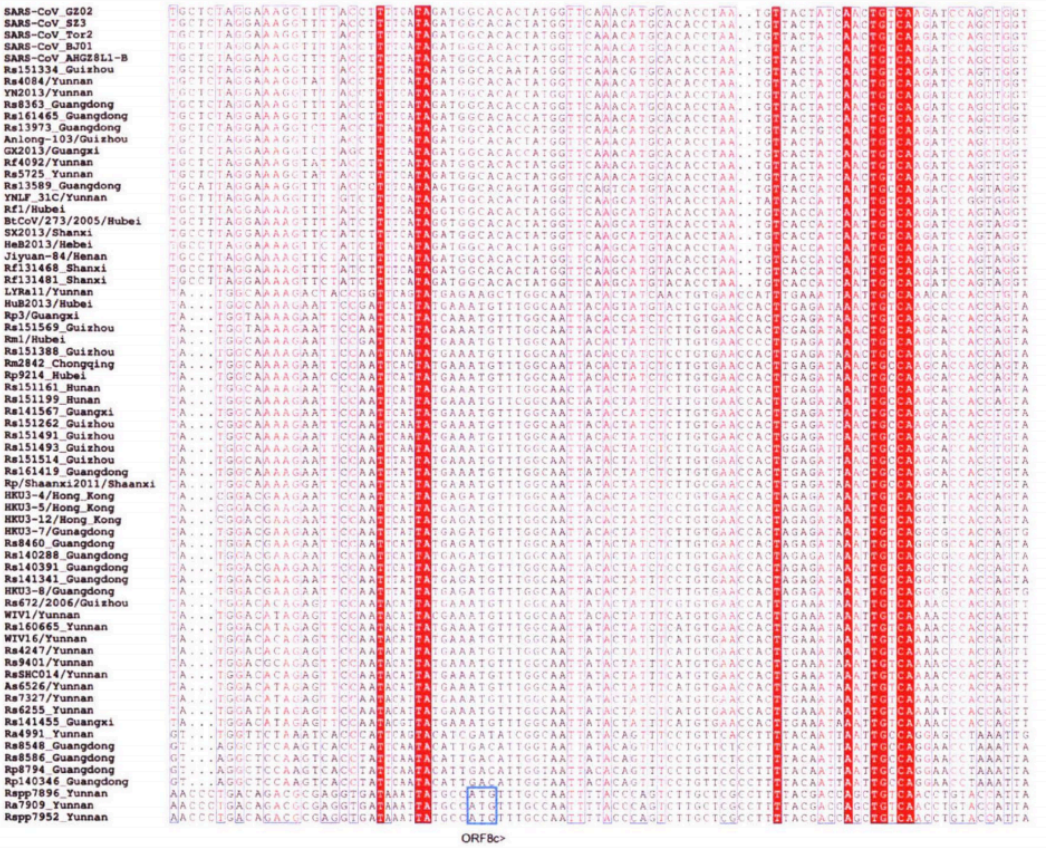
图 3.4 ORF8 不同基因结构的示意图。

ORF8 中核苷酸的缺失用蓝色虚线框表示，虚线框内部的数字为缺失的核苷酸数量。

Figure 3.4 Schematic diagram of different gene structures of ORF8.


ORF8/ORF8a>
ORF8b>
SARS-CoV_GZ02
SARS-CoV_SZ3
SARS-CoV_Tor2
SARS-CoV_BJ01
SARS-CoV_AHGZ8L1-B
Rs151334_Guizhou
Rs4084/Yunnan
YN2013/Yunnan
Rs8363_Guangdong
Rs161465_Guangdong
Rs13973_Guangdong
Anlong-103/Guizhou
GX2013/Guangxi
Rf4092/Yunnan
Rs5725_Yunnan
Rs13589_Guangdong
YNLF_31C/Yunnan
Rf1/Hubei
BtCoV/273/2005/Hubei
SX2013/Shanxi
HeB2013/Hebei
Jiyuan-84/Henan
Rf131468_Shanxi
Rf131481_Shanxi
LYRa11/Yunnan
HuB2013/Hubei
Rp3/Guangxi
Rs151569_Guizhou
Rm1/Hubei
Rs151388_Guizhou
Rm2842_Chongqing
Rp9214_Hubei
Rs151161_Hunan
Rs151199_Hunan
Rs141567_Guangxi
Rs151262_Guizhou
Rs151491_Guizhou
Rs151493_Guizhou
Rs151514_Guizhou
Rs161419_Guangdong
Rp/Shaanxi2011/Shaanxi
HKU3-4/Hong_Kong
HKU3-5/Hong_Kong
HKU3-12/Hong_Kong
HKU3-7/Gunagdong
Rs8460_Guangdong
Rs140288_Guangdong
Rs140391_Guangdong
Rs141341_Guangdong
HKU3-8/Guangdong
Rs672/2006/Guizhou
WIV1/Yunnan
Rs160665_Yunnan
WIV16/Yunnan
Rs4247/Yunnan
Rs9401/Yunnan
RsSHC014/Yunnan
As6526/Yunnan
Rs7327/Yunnan
Rs6255_Yunnan
Rs141455_Guangxi
Ra4991_Yunnan
Rs8548_Guangdong
Rs8586_Guangdong
Rp8794_Guangdong
Rp140346_Guangdong
Rspp7896_Yunnan
Ra7909_Yunnan
Rspp7952_Yunnan
ORF8a>
ORF8b>


ORF8a<
SARS-CoV_GZ02
SARS-CoV_SZ3
SARS-CoV_Tor2
SARS-CoV_BJ01
SARS-CoV_AHGZ8L1-B
Rs151334_Guizhou
Rs4084/Yunnan
YN2013/Yunnan
Rs8363_Guangdong
Rs161465_Guangdong
Rs13973_Guangdong
Anlong-103/Guizhou
GX2013/Guangxi
Rf4092/Yunnan
Rs5725_Yunnan
Rs13589_Guangdong
YNLF_31C/Yunnan
Rf1/Hubei
BtCoV/273/2005/Hubei
SX2013/Shanxi
HeB2013/Hebei
Jiyuan-84/Henan
Rf131468_Shanxi
Rf131481_Shanxi
LYRa11/Yunnan
HuB2013/Hubei
Rp3/Guangxi
Rs151569_Guizhou
Rm1/Hubei
Rs151388_Guizhou
Rm2842_Chongqing
Rp9214_Hubei
Rs151161_Hunan
Rs151199_Hunan
Rs141567_Guangxi
Rs151262_Guizhou
Rs151491_Guizhou
Rs151493_Guizhou
Rs151514_Guizhou
Rs161419_Guangdong
Rp/Shaanxi2011/Shaanxi
HKU3-4/Hong_Kong
HKU3-5/Hong_Kong
HKU3-12/Hong_Kong
HKU3-7/Gunagdong
Rs8460_Guangdong
Rs140288_Guangdong
Rs140391_Guangdong
Rs141341_Guangdong
HKU3-8/Guangdong
Rs672/2006/Guizhou
WIV1/Yunnan
Rs160665_Yunnan
WIV16/Yunnan
Rs4247/Yunnan
Rs9401/Yunnan
RsSHC014/Yunnan
As6526/Yunnan
Rs7327/Yunnan
Rs6255_Yunnan
Rs141455_Guangxi
Ra4991_Yunnan
Rs8548_Guangdong
Rs8586_Guangdong
Rp8794_Guangdong
Rp140346_Guangdong
Rspp7896_Yunnan
Ra7909_Yunnan
Rspp7952_Yunnan
ORF8a<
ORF8b<


SARS-CoV_GZ02
SARS-CoV_SZ3
SARS-CoV_Tor2
SARS-CoV_BJ01
SARS-CoV_AHGZ8L1-B
Rs151334_Guizhou
Rs4084/Yunnan
YN2013/Yunnan
Rs8363_Guangdong
Rs161465_Guangdong
Rs13973_Guangdong
Anlong-103/Guizhou
GX2013/Guangxi
Rf4092/Yunnan
Rs5725_Yunnan
Rs13589_Guangdong
YNLF_31C/Yunnan
Rf1/Hubei
BtCoV/273/2005/Hubei
SX2013/Shanxi
HeB2013/Hebei
Jiyuan-84/Henan
Rf131468_Shanxi
Rf131481_Shanxi
LYRa11/Yunnan
HuB2013/Hubei
Rp3/Guangxi
Rs151569_Guizhou
Rm1/Hubei
Rs151388_Guizhou
Rm2842_Chongqing
Rp9214_Hubei
Rs151161_Hunan
Rs151199_Hunan
Rs141567_Guangxi
Rs151262_Guizhou
Rs151491_Guizhou
Rs151493_Guizhou
Rs151514_Guizhou
Rs161419_Guangdong
Rp/Shaanxi2011/Shaanxi
HKU3-4/Hong_Kong
HKU3-5/Hong_Kong
HKU3-12/Hong_Kong
HKU3-7/Gunagdong
Rs8460_Guangdong
Rs140288_Guangdong
Rs140391_Guangdong
Rs141341_Guangdong
HKU3-8/Guangdong
Rs672/2006/Guizhou
WIV1/Yunnan
Rs160665_Yunnan
WIV16/Yunnan
Rs4247/Yunnan
Rs9401/Yunnan
RsSHC014/Yunnan
As6526/Yunnan
Rs7327/Yunnan
Rs6255_Yunnan
Rs141455_Guangxi
Ra4991_Yunnan
Rs8548_Guangdong
Rs8586_Guangdong
Rp8794_Guangdong
Rp140346_Guangdong
Rspp7896_Yunnan
Ra7909_Yunnan
Rspp7952_Yunnan
ORF8c>


ORF8b<
SARS-CoV_GZ02
SARS-CoV_SZ3
SARS-CoV_Tor2
SARS-CoV_BJ01
SARS-CoV_AHGZ8L1-B
Rs151334_Guizhou
Rs4084/Yunnan
YN2013/Yunnan
Rs8363_Guangdong
Rs161465_Guangdong
Rs13973_Guangdong
Anlong-103/Guizhou
GX2013/Guangxi
Rf4092/Yunnan
Rs5725_Yunnan
Rs13589_Guangdong
YNLF_31C/Yunnan
Rf1/Hubei
BtCoV/273/2005/Hubei
SX2013/Shanxi
HeB2013/Hebei
Jiyuan-84/Henan
Rf131468_Shanxi
Rf131481_Shanxi
LYRa11/Yunnan
HuB2013/Hubei
Rp3/Guangxi
Rs151569_Guizhou
Rm1/Hubei
Rs151388_Guizhou
Rm2842_Chongqing
Rp9214_Hubei
Rs151161_Hunan
Rs151199_Hunan
Rs141567_Guangxi
Rs151262_Guizhou
Rs151491_Guizhou
Rs151493_Guizhou
Rs151514_Guizhou
Rs161419_Guangdong
Rp/Shaanxi2011/Shaanxi
HKU3-4/Hong_Kong
HKU3-5/Hong_Kong
HKU3-12/Hong_Kong
HKU3-7/Gunagdong
Rs8460_Guangdong
Rs140288_Guangdong
Rs140391_Guangdong
Rs141341_Guangdong
HKU3-8/Guangdong
Rs672/2006/Guizhou
WIV1/Yunnan
Rs160665_Yunnan
WIV16/Yunnan
Rs4247/Yunnan
Rs9401/Yunnan
RsSHC014/Yunnan
As6526/Yunnan
Rs7327/Yunnan
Rs6255_Yunnan
Rs141455_Guangxi
Ra4991_Yunnan
Rs8548_Guangdong
Rs8586_Guangdong
Rp8794_Guangdong
Rp140346_Guangdong
Rspp7896_Yunnan
Ra7909_Yunnan
Rspp7952_Yunnan
ORF8c<

Figure 3.5 Nucleotide sequence comparison results of ORF8.

Figure 3.5 Nucleotide sequence alignment of ORF8.

Figure 3.6 Phylogenetic tree built based on ORF8b.

Virulent strains with deletions in ORF8 are indicated by red asterisks. SARS-like coronaviruses found in this study are marked in bold.

Figure 3.6 The phylogeny of ORF8b.

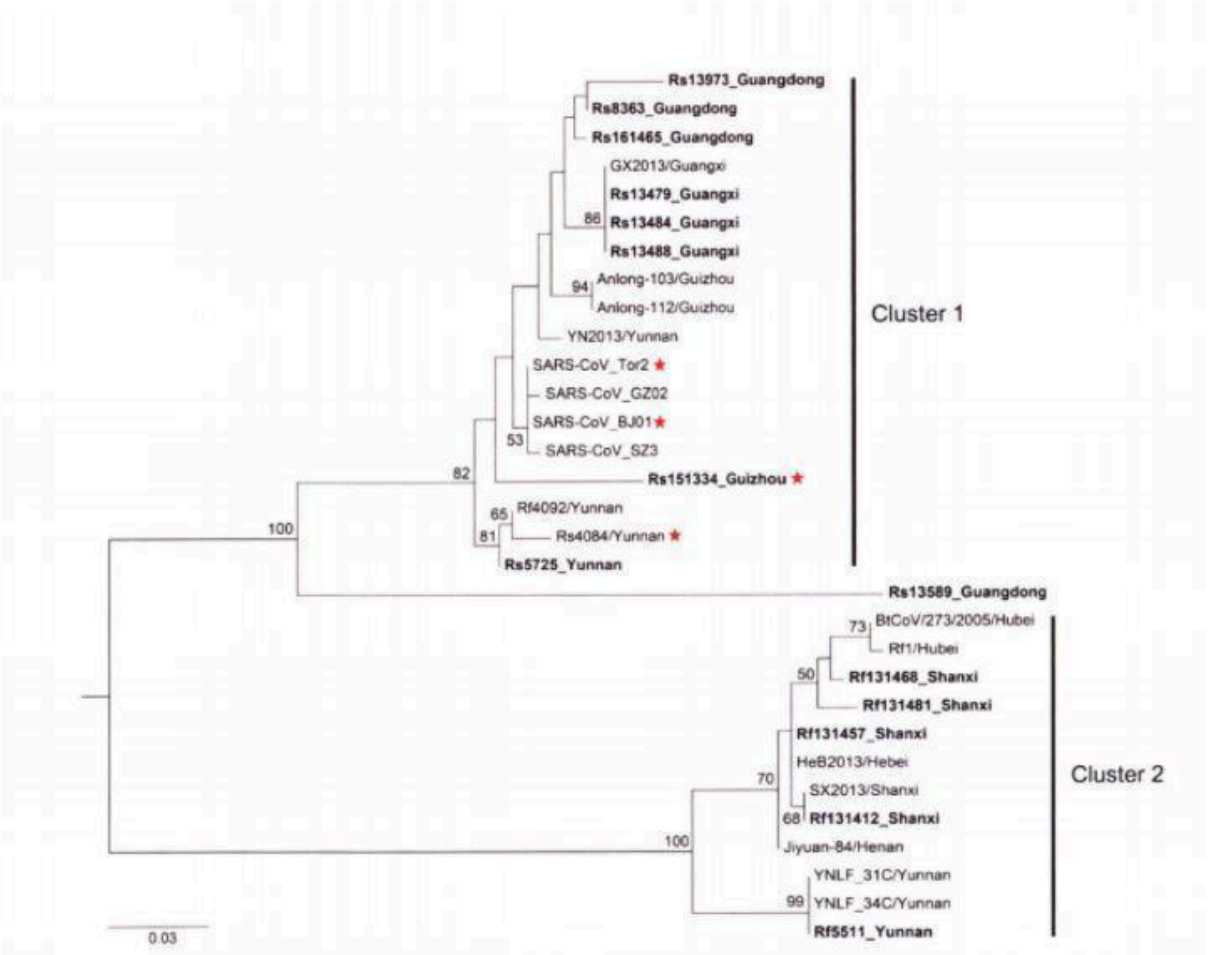


图 3.6 基于 ORF8b 建立的系统进化树。

ORF8 有缺失的毒株用红色的星号表示。本研究发现的 SARS 样冠状病毒标粗。

Figure 3.6 The phylogeny of ORF8b.

The strains with deletions in ORF8 genes are marked with the red pentagram.The SARSr-CoV strains discovered in this study are shown in bold.

3.6 Coevolutionary analysis of SARS-like coronaviruses and hosts

The evolution and development of viruses is not only related to the phylogeny of the virus itself, but also to host interactions. Bats are the only flying mammals, and intercommunication between bats can facilitate intercommunication between viruses. To investigate the relationship between virus and host, coevolutionary analysis was performed using the full-length RdRp sequence of the virus and CytB of the host. From Figure 3.7, SARS-like coronavirus evolves mainly in homozygous bat species, followed by host switching in different species of hosts, indicating that the host has a limiting effect on the evolution of SARS-like coronavirus. To further verify the co-evolutionary relationship between virus and host, the evolutionary tree constructed by the full-length RdRp sequence of the virus and the evolutionary tree constructed by the CytB sequence of the host were combined, and it was found that SARS-like coronaviruses with close relatives in the same region such as Zhejiang existed in different bat species, indicating that SARS-like coronaviruses lacked strict host restriction. Host switching of SARS-like coronaviruses occurred mainly in different bat species of the family Chrysomelidae, indicating that host genetic distance is an important factor influencing the occurrence of host switching and cross-species transmission of viruses. SARS-like coronaviruses from the same bat species such as the big-eared chrysanthemum-headed bat (R. macrotis) were found to be more distantly related, suggesting that the virus evolved in a homogeneous host. In addition, SARS-like coronaviruses from the same species such as R. ferrumequinum tend to cluster according to geographically proximate locations, suggesting that SARS-like coronaviruses are more geographically restricted than host-restricted.

Figure 3.7 Estimation of co-phylogenetic events of SARS-CoVs, SARSr-CoVs and their hosts.

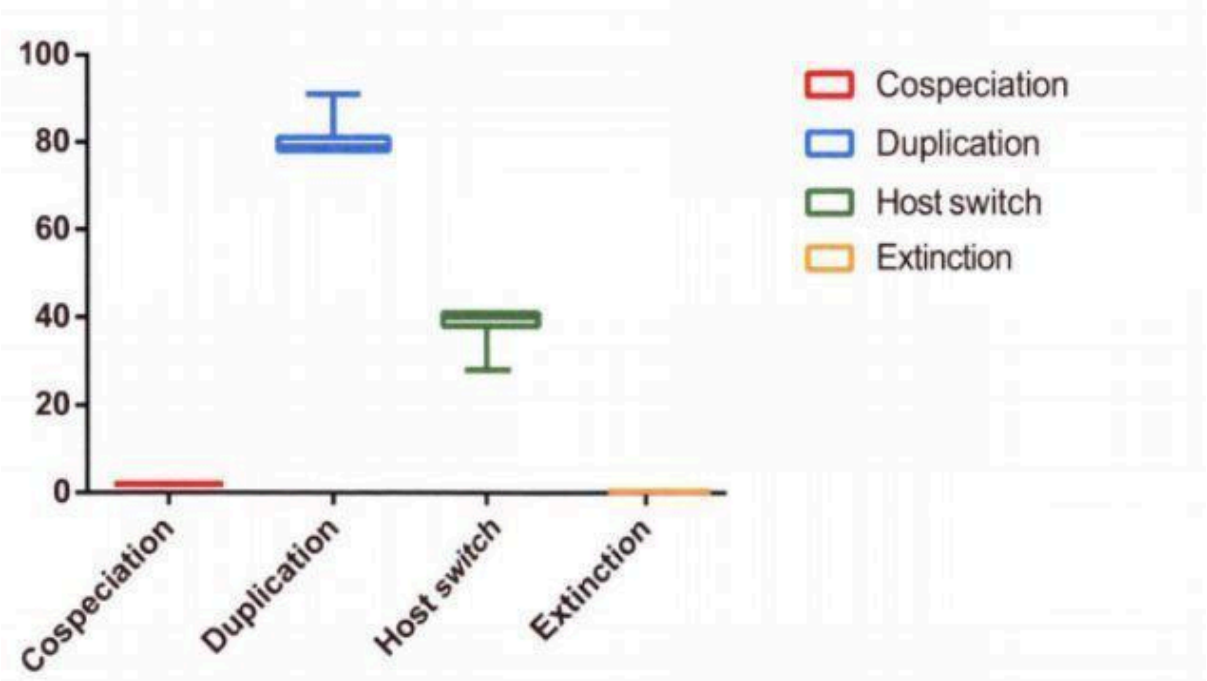


图 3.7 SARS 冠状病毒和 SARS 样冠状病毒与其宿主共同发育事件的估计。

Figure 3.7 Estimation of co-phylogenetic events of SARS-CoVs, SARSr-CoVs and their hosts.

Figure 3.8 Co-phylogenetic tree of SARS coronavirus and SARS-like coronavirus with their hosts.

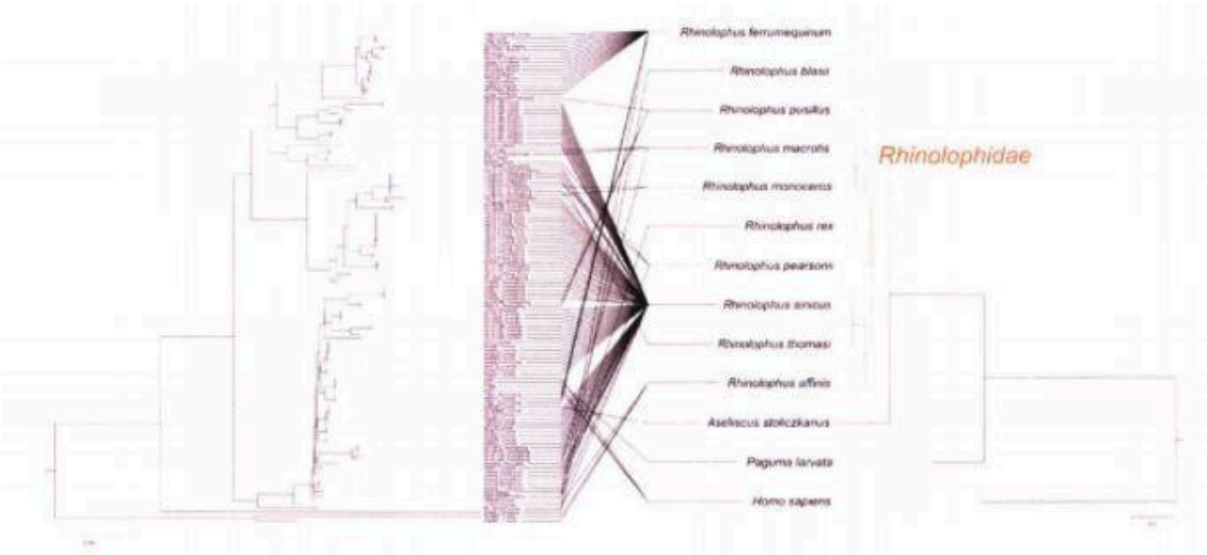


图 3.8 SARS 冠状病毒和 SARS 样冠状病毒与其宿主的共系统进化发育树。

The branches of the family Chrysomelidae are marked in orange, and at the nodes are marked as Rhinolophidae.

Figure 3.8 Cophylogeny of SARS-CoVs, SARSr-CoVs and their hosts.

The branch of Rhinolophidae is shown in orange and named Rhinolophidae.

3.7 Discussion

Currently, bat SARS-like coronaviruses have been identified in countries in Asia, Europe, and Africa, however, among non-Chinese countries only one strain of SARS-like coronavirus with full-length genome sequence was obtained from Korea and Bulgaria, respectively, and most other countries only have a small SARS-like coronavirus gene sequence for coronal detection, and an evolutionary tree constructed based on this short sequence showed that SARS-like coronaviruses from China are more closely related to human SARS coronaviruses than other countries [71-81]. In this study, a large-scale SARS-like coronavirus screen of bats SARS-like coronaviruses collected in twenty regions of China from 2011-2016 available in the laboratory showed that positive SARS-like coronaviruses were detected mainly in bats of the family Chrysomelidae, and high-throughput sequencing was performed on four of these strains, and only the amino acid sequences of individual genes of Rs5725_Yunnan Rs5725 was relatively high with that of human SARS coronavirus.

The RdRp phylogenetic analysis of SARS-like coronaviruses showed that SARS-like coronaviruses tended to cluster together in adjacent areas or in the same area, indicating that overlapping host geographic locations played an important role in the evolution of SARS-like coronaviruses, i.e., geographic restriction. Marker recapture experiments demonstrated that the maximum migration distance of Chinese chrysanthemum-headed bats is 17 km, and other bat species of chrysanthemum-headed bats can migrate up to 30 km for hibernation, and the migration of bats contributes to the spread of viruses carried by bats over a certain geographical range [73]. The evolutionary tree can be divided into four main lineages, and SARS-like coronaviruses from lineage 1 (Yunnan, Guangxi, Guizhou, Guangdong, and Chongqing) are closest to SARS coronaviruses, indicating that human SARS coronaviruses may have originated in these regions. These regions are geographically proximate and bats can fly, and there is a possibility that bats from different regions communicate with each other and thus promote virus intercommunication. Also, the species diversity of SARS-like coronaviruses in the southern regions is higher than that in the northern regions, which may be related to the higher genetic diversity of bats in the southern regions than in the northern regions.

The highly variable S gene and ORF8 gene of SARS-like coronavirus were analyzed. Compared with the RBM region of SARS coronaviruses, SARS-like coronaviruses without deletions in this region are currently found only in Yunnan, and these viruses belong to the same cluster as SARS-like coronavirus WIV1, which was previously reported to be available for bat, civet, and human ACE2 receptors, indicating that the S gene of human SARS coronaviruses may have originated in Yunnan. In addition, SARS-like coronaviruses with deletions in the RBM region were found in Yunnan, indicating that the diversity of SARS-like coronaviruses in Yunnan is relatively high [90]. Recombination analysis of S genes showed frequent recombination of S genes in bat SARS-like coronaviruses. Gene matching analysis of the ORF8 gene of SARS-like coronaviruses showed that the vast majority of SARS-like coronaviruses had no deletion of ORF8. ORF8-deficient SARS-like coronaviruses were found in Guizhou and Yunnan in this study, in addition to the previously reported regions of Yunnan, Guangdong and Bulgaria, and the number of missing bases was different, showing the genetic diversity of ORF8 of SARS-like coronaviruses. However, in addition to complete

deletion, the same deletion as in the middle and late stages of human SARS coronaviruses was not found, indicating that the evolution of this gene is complex. Phylogenetic analysis of ORF8b showed that SARS coronavirus is most closely related to those from cluster 1 (Yunnan, Guizhou, Guangxi and Guangdong), further demonstrating that SARS coronavirus originated from SARS-like coronavirus in the lineage 1 region.

Coevolutionary analysis of SARS-like coronaviruses and their hosts showed that the viruses evolved mainly in the homologous host, the Chinese chrysanthemum-headed bat, suggesting that hosts play a limiting role in virus evolution, which explains why most SARS coronaviruses are found in the Chinese chrysanthemum-headed bat, but this limiting role is not strict, and host switching occurs in different bat species of the family Chrysomelidae, which may be due to the fact that closely related hosts provide a similar environment for the virus to survive.

By analyzing the three specific genes of SARS-like coronavirus and the relationship between the virus and the host, this study is important for elucidating the geographical origin of SARS coronavirus and the evolution of SARS-like coronavirus in bats, and providing a theoretical basis for subsequent prevention and control.

Chapter 4 Conclusion and Outlook

4.1 Conclusion

1.This study investigated the prevalence of SARS-like coronaviruses in twenty regions of China during 2011-2016 and detected positively in nine regions, among which, Chongqing, Hunan and Beijing were the first time SARS-like coronaviruses were found. High-throughput sequencing of four of these SARS-like coronaviruses was performed, and genomic analysis showed that except for ORF8, all other genes of Rs5725_Yunnan, including S and ORF3, had high amino acid similarity with SARS coronavirus.

2. Phylogenetic analyses all showed that the bat SARS-like coronaviruses with the closest affinity to the RdRp gene, S gene and ORF8 gene of SARS coronavirus were all present in Yunnan, a finding consistent with previous literature reports that similar viruses circulate in southern China (Yunnan, Guangxi, Guizhou, Guangdong and Chongqing) and, at the same time, these regions are proximate, indicating geographical overlap has an important role in the evolution and exchange of SARS-like coronaviruses [87]. In addition, bat SARS-like coronaviruses tend to cluster together more by geographic location than by host species, suggesting that space plays a greater barrier role.

3. Relative to SARS coronaviruses, bat SARS-like coronaviruses without deletions in the RBM region were only found in Yunnan, while SARS-like coronaviruses with deletions in ORF8 were found not only in Yunnan but also in Guangdong, Guizhou, and Bulgaria, indicating that the S gene of SARS coronaviruses originated in Yunnan, and also suggesting that ORF8 is less geographically restricted than the S gene, except for complete deletion, and no ORF8 deletion was found as in the middle and late stages of human SARS coronavirus.

4. There is widespread recombination in the S gene of SARS-like coronaviruses, and recombination loci occur mainly in SARS-like coronaviruses of family 1 and family 2, and these recombinations occur in both the same and different regions, and the presence of recombination enriches the species diversity of SARS-like coronaviruses.

5. SARS-like coronaviruses in bats evolved and mutated mainly in bats of the same species of the family Rhinolophus, but host switching also existed and could occur in bats of different species of the same family of Rhinolophus, or even in bats of different families such as Rhinolophus and Hipposideros, indicating that SARS-like coronaviruses exchanged in bats of the same or different species, and in comparison, hosts with close relatives were conducive to host switching.

4.2 Problems and outlook

1. Although SARS coronavirus was first outbreak in Guangdong, the relatives of bat SARS-like coronavirus and human SARS coronavirus found in Guangdong are not very close at present. Previous literature as well as the present study suggest that human SARS coronavirus likely originated from SARS-like coronavirus in Yunnan. Although bats of the family Rhinolophus can migrate up to 30 km, such a long distance from Yunnan to Guangdong seems unlikely, and there may be a role of intermediate hosts as well as human activities. Subsequent microsatellite experiments of hosts can be used to trace the gene flow of hosts to investigate the mode and mechanism of SARS-like coronavirus transmission and migration of bats to Guangdong. In addition, the population infection in the area of circulation of bat SARS-like coronavirus should be continuously monitored to provide assistance for the prevention and control of this type of virus.

2. In contrast to the RBM region of SARS coronaviruses, the vast majority of bat SARS-like coronaviruses identified so far have deletions in the RBM region, and these viruses have been shown to be unable to utilize the host ACE receptor, and it is unclear whether these bat SARS-like coronaviruses that cannot utilize the ACE receptor have other receptors to help them invade host cells.

3. ORF8 of SARS-like coronaviruses is not essential for virus replication, and the ORF8-deficient SARS-like coronaviruses are not confined to one region, but are scattered in several regions. Meanwhile, in addition to complete deletion, the same deletion as that of human SARS coronaviruses in the middle or late stages is not found for the time being, so the evolution of ORF8 of SARS coronaviruses still needs further study.

4. The bat SARS-like coronavirus mainly spreads and evolves within or between species in the family Rhinolophus, and it is unclear whether it is related to the species specificity and immunity of Rhinolophus, so the study of the host is also an urgent issue for future research work, which is an important guideline for the study of virus transmission.

References

[1] DE GROOT R J, BAKER S C, BARIC R, et al. Coronaviridae[J]. Virus Taxonomy：Ninth Report of the International Committee on Taxonomy ofViruses, 20 1 1：806—28.
[2] SCHALK A F, HAWN M C. An apparently new respiratory disease of baby chicks[J]. Jamvetmedassoc, 1 93 1, 78：41 3-422.
[3] PFEFFERLE S, OPPONG S, DREXLER J F, et al. Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana[J]. Emerging Infectious Diseases, 2009, 15(9)：1 377-84.
[4] HUYNH J, LI S, YOUNT B, et al. Evidence supporting a zoonotic origin of human coronavirus strain NL63[J]. Journal ofVirology,20 1 2, 86(23)：128 1 6-25.
[5] GUAN Y ZHENG B J, HE Y Q, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China[J]. Science(New York, NY), 2003, 302(5643)： 276. 8.
[6] KSIAZEK T G, ERDMAN D, GOLDSMITH C S, et al. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome[J]. New England Journal of Medicine, 2003, 348(20)： 1953. 66.
[7] DROSTEN C, GUNTHER S, PREISER W：et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome[J]. The New England Journal of Medicine, 2003, 348(20)：1967-76.
[8] PEIRIS J S M, GUAN YYUEN K Y Severe acute respiratory syndrome[J]. Nature Medicine, 2004, 1 0(12 Suppl)：$88-97.
[9] ZHONG N S, ZHENG B J, LI Y M, et al. Epidemiology and cause of severe acute respiratory syndrome(SARS)in Guangdong, People'S Republic of China, in February,2003[J]. Lancet(London, England), 2003, 362(9393)：1 353-8.
[10] ZAKI A M, VAN BOHEEMEN S, BESTEBROER TM, et al. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia[J]. The New England Journal of Medicine, 20 1 2, 367(19)：1814-20.
[11] PAYNE S. Chapter 17-FamilyCoronaviride[M] . Viruses. Academic Press. 2017：149—58
[12] CUI J, LI F'SHI Z L. Origin and evolution of pathogenic coronaviruses[J]. Nature Reviews Microbiology,20 1 9, 1 7(3)：I 8 1-92.
[13] WOO P C, LAU S K, LAM C S, et al. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus[J]. Journal ofVirology,2012, 86(7)：3995·4008.
[14] RAJ V S, STERHAUS A D, FOUCHIER R A, et al. MERS：emergence of a novel human coronavirus[J]. Current Opinion in Virology,20 1 4, 5：58-62
[15] SU S, WONG G, SHl wj et al. Epidemiology,Genetic Recombination, and Pathogenesis o fCoronaviruses[J]. Trends in Microbiology,20 16, 24(6)：490—502
[16] WOO P C, WANG M, LAU S K et al. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features[J]. Journal of Virology,2007, 81(4)：1574-85.
[17] ANTHONY S J, JOHNSON CK GREIG D J, et al. Global patterns in coronavirus diversity[J] t Virus Evol, 2017, 3(1)：vex012.
[18] DREXLER J F, CORMAN V M, DROSTEN C. Ecology,evolution and classification of bat coronaviruses in the aftermath of SARS[J]. Antiviral Research, 20 1 4, 1 0 1：45—56
[19] FAN Y ZHAO K, SHI Z L, et al. Bat Coronaviruses in China[J]. Viruses, 2019, 1 1(3)

[20] AZHAR E I, EL—KAFRAWY S A, FARRAJ S A, et al. Evidence for camel to—human transmission of MERS coronavirus[J]. The New England Journal of Medicine, 2014, 370(26) 2499—505.
[21] WANG Y LIU D, SHI Wj et al. Origin and Possible Genetic Recombination of the Middle East Respiratory Syndrome Coronavirus from the First ImpoSed Case in China：Phylogenetics and Coalescence Analysis[J]. mBio, 2015, 6(5)：e01280—15.
[22] DU L, HAN G z. Deciphering MERS—CoV Evolution in Dromedary Camels[J]. Trends in Microbiology,20 1 6, 24(2)：87—9
[23] MACKAY I M, ARDEN K E. MERS coronavirus：diagnostics, epidemiology and transmission [J]. Virology Journal, 2015, 12：222
[24] ZHOU P, FAN H, LAN T, et al. Fatal swine acute diarrhoea syndrome caused by an HKU2一related coronavirus of bat origin[J]. Nature, 2018, 556(7700)：255—8
[25] BALLESTEROS M L, SANCHEZ C M, ENJUANES L. Two amino acid changes at the N—terminus of transmissible gastroenteritis coronavirus spike protein result in the loss of enteric tropism[J]. Virology,1997, 227(2)：378-88.
[26] KREMPL C, SCHULTZE B, LAUDE H, et al. Point mutations in the S protein connect the sialic acid binding activity with the enteropathogenicity of transmissible gastroenteritis coronavirus[J]. Journal ofVirology,1997, 71(4)：3285-7.
[27] SCHICKLI J H, THACKRAY L B, SAWICKI S G, et al. The N—terminal region of the murine coronavirus spike glycoprotein is associated with the extended host range of viruses from persistently infected murine cells[J]. Journal ofVirology,2004, 78(17)：9073—83.
[28] DE HAAN C A, TE LINTELO E, LI Z, et al. Cooperative involvement of the S 1 and S2 subunits of the murine coronavirus spike protein in receptor binding and extended host range[J]. Journal ofVirology,2006, 80(22)：10909—18
[29] TUSELL S M, SCHITTONE S A, HOLMES K V Mutational analysis of aminopeptidase N, a receptor for several group 1 coronaviruses, identifies key determinants of viral host range[J]. Journal ofVirology,2007, 8 1(3)：1261-73.
[30] LI F. Receptor recognition and cross—species infections of SARS coronavirus[J]. Antiviral Research, 2013, 100(1)：246—54.
[31] ENJUANES L, ALMAZAN F, SOLA I, et al. Biochemical aspects of coronavirus replication[J] Advances in Experimental Medicine and Biology,2006, 58 1：13—24.
[32] HEALD—SARGENT T, GALLAGHER T. Ready,set, fuse!The coronavirus spike protein and acquisition offusion competence[J]. Viruses, 2012, 4(4)：557—80
[33] Han Bai Ela, He Weiping. Advances in the mechanism of SARS coronavirus cell invasion[J]. Infectious Disease Information, 2006, 19(1):23-5.
[34] RUCH T R, MACHAMER C E. The coronavirus E protein：assembly and beyond[J] _Viruses, 2012, 4(3)：363-82.
[35] ESCORS D, ORTEGO J, LAUDE H, et al. The membrane Mprotein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability[J]. Journal of Virology, 2001, 75(3)：13 12—24.
[36] SIU K L, CHAN C P'KOK K H, et al. Suppression of innate antiviral response by severe acute respiratory syndrome coronavirus Mprotein is mediated through the first transmembrane domain[J]. Cellular&Molecular Immunology,2014, I l(2)：141-9.
[37] NEUMAN B W：KISS G, KUNDING A H, et al. A structural analysis of M protein in coronavirus assemblyandmorphology[J]. Journal ofStructuralBiology,2011, 174(1)：1l·22.
[38] ROTTIER P J M. The molecular dynamics of feline coronaviruses[J]. Veterinary Microbiology, 1999, 69(1)：1 17—25.

[39] KLAUSEGGER A, STROBL B, REGL G, et al. Identification of a coronavirus hemagglutinin-esterase with a substrate specificity different from those of influenza C virus and bovine coronavirus[J]. Journal of Virology, 1999, 73(5):3737-43.
[40] CORNELISSEN L A, WIERDA C M, VAN DER MEER F J, et al. Hemagglutinin-esterase, a novel structural protein of torovirus[J]. Journal ofVirology,1997, 7l(7):5277·86.
[41] KAZI L, LISSENBERG A, WATSON R, et al. Expression of hemagglutinin esterase protein from recombinant mouse hepatitis virus enhances neurovirulence[J]. Journal of Virology,2005, 79(24):1 5064-73.
[42] CHANG C K SUE S C, YU T H, et al. Modular organization of SARS coronavirus nucleocapsid protein[J]. Journal ofBiomedical Science, 2006, 13(1):59-72.
[43] HURST K R, KOETZNER C A, MASTERS P S. Identification of in vivo-interacting domains of the murine coronavirus nucleocapsid protein[J]. Journal ofVirology,2009, 83(14):7221·34.
[44] STOHLMAN S A, LAJ M M. Phosphoproteins of murine hepatitis viruses[J]. Journal of Virology,1 979, 32(2):672—5.
[45] FEHR A R, PERLMAN S. Coronavimses:an overview of their replication and pathogenesis[J]. Methods in Molecular Biology(Clifton, NJ), 2015, 1282:1-23.
[46] BARANOV P V HENDERSON C M, ANDERSON C B, et al. Programmed ribosomal frameshiRing in decoding the SARS-CoV genome[J]. Virology,2005, 332(2):498-5 1 0.
[47] BRIERLEY I, DIGARD P, INGLIS S C. Characterization of an efficient coronavirus ribosomal frameshifting signal:requirement for an RNA pseudoknot[J]. Cell, 1989, 57(4):537-47.
[48] DE WIT E, VAN DOREMALEN N, FALZARANO D, et al. SARS and MERS:recent insights into emerging coronaviruses[J]. Nature Reviews Microbiology,2016, 14(8):523-34.
[49] PERLMAN S, NETLAND J. Coronaviruses post-SARS:update on replication and pathogenesis [J]. Nature Reviews Microbiology,2009, 7(6):439-50.
[50] VEGA V B, RUAN Y, LIU J, et al. Mutational dynamics of the SARS coronavirus in cell culture and human populations isolated in 2003[J]. BMC infectious diseases, 2004, 4:32.
[51] LIAO C L, LAI M M. RNA recombination in a coronavirus:recombination between viral genomic RNAandtransfected RNAfragments[J]. Journal ofVirology,1992, 66(10):6117—24.
[52] NAVAS-MARTIN S, BROM M, CHUA M M, et al. Replicase genes of murine coronavirus strains A59 and JHM are interchangeable:differences in pathogenesis map to the 3'one—third of the genome[J]. Journal of Virology, 2007, 81(2):1 022-6.
[53] LAU S K, LEE P, TSANG A K, et al. 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]. Journal ofVirology,201 1, 85(21):1 1325-37.
[54] PYRC K, DIJKMAN R, DENG L, et al. Mosaic structure of human coronavirus NL63, one thousand years of evolution[J]. Journal of Molecular Biology,2006, 364(5):964-73.
[55] HUANG C, LIU W J, XU W:et al. A Bat-Derived Putative Cross-Family Recombinant Coronavims with a Reovirus Gene[J]. PLoS pathogens, 2016, 12(9):e1005883.
[56] PARRISH C 1L HOLMES E C, MORENS D M, et al. Cross-species virus transmission and the emergence of new epidemic diseases[J]. Microbiology and Molecular Biology Reviews:MMBR, 2008, 72(3):457-70.
[57] RAJ V S, MOU H, SMITS S L, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus·EMC[J]. Nature, 2013, 495:251.

[58] MULLER M A, RAJ V S, MUTH D, et al. Human coronavirus EMC does not require the SARS-coronavirus receptor and maintains broad replicative capability in mammalian cell lines[J]. mBio, 2012, 3(6).
[59] CAVANAGH D, MAWDITT K, ADZHAR A, et al. Does IBV change slowly despite the capacity of the spike protein to vary greatly?[J]. Advances in Experimental Medicine and Biology, 1998, 440：729-34.
[60] WESLEY R D. The S gene of canine coronavirus, strain UCD-1, is more closely related to the S gene of transmissible gastroenteritis virus than to that of feline infectious peritonitis virus[J]. Virus Research, 1 999, 6 1(2)：145-52.
[61] WANG L, JUNKER D, COLLISSON E W：Evidence of natural recombination within the S 1 gene of infectious bronchitis virus[J]. Virology,1 993, 1 92(2)：7 1 0-6.
[62] JIA W：WANG X, PARRISH C R, et al. Analysis of the serotype-specific epitopes of avian infectious bronchitis virus strains Ark99 and Mass41[J]. Journal ofVirology,1996, 70(10)：7255·9.
[63] Wang F S, Xu D P. Characterization and pathogenesis of SARS coronavirus in Sichuan. Infectious Disease Information, 2003: 67. 8.
[64] TU C, CRAMERI G, KONG X, et al. Antibodies to SARS coronavirus in civets[J]. Emerging Infectious Diseases, 2004, 1 0(1 2)：2244-8.
[65] KAN B, WANG M, JING H, et al. Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms[J]. Journal ofVirology,2005, 79(18)：1 1892-900.
[66] SONG H D, TU C C, ZHANG G W, et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human[J]. Proceedings of the National Academy of Sciences of the United States ofAmerica, 2005, 102(7)：2430-5.
[67] SHI Z, HU Z. A review of studies OB animal reservoirs of the SARS coronavirus[J]. Virus Research, 2008, 1 33(1)：74-87.
[68] CHAN P K, CHAN M C. Tracing the SARS·coronavirus[J]. Journal of Thoracic Disease, 20 1 3, 5 Suppl 2：S118—21.
[69] LAU S K, WOO P C, LI K S, et al. Severe acute respiratory syndrome coronavirus·like virus in Chinese horseshoe bats[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(39)：14040-5.
[70] LI W：SHI Z, YU M, et ai. Bats are natural reservoirs of SARS-like coronaviruses[J]. Science (New York, NV), 2005, 3 1 0(5748)：676·9.
[71] REN Wj LI Wj YU M, et al. Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis[J]. The Journal of General Virology,2006, 87(Pt 1 1)： 3355-9.
[72] DREXLER J F, GLOZA—RAUSCH F, GLENDE J, et al. Genomic characterization of severe acute respiratory syndrome·related coronavirus in European bats and classification of coronaviruses based On partial RNA-dependent RNA polymerase gene sequences[J]. Journal of Virology,20 1 0, 84(21)：11336-49.
[73] LAU S K, LI K S, HUANG Y et al. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events[J]. Journal ofVirology,2010, 84(6)：2808-19.
[74] QUAN P L, FIRTH C, STREET C, et al. Identification of a severe acute respiratory syndrome coronavirus-like virus in a leaf-nosed bat in Nigeria[J]. mBio, 2010, 1(4).

[75] RIHTARIC D, HOSTNIK P, STEYER A, et al. Identification of SARS-like coronaviruses in horseshoe bats(Rhinolophus hipposideros)in Slovenia[J]. Archives of Virology,20 1 0, 1 55(4) : 507. 14.
[76] YUAN J, HON C C, LI Y et al. Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans[J]. The Journal of General Virology,20 1 0, 9 1(Pt 4):1058-62.
[77] BALBONI A, GALLINA L, PALLADINI A, et al. A real-time PCR assay for bat SARS-like coronavirus detection and its application to Italian greater horseshoe bat faecal sample surveys[J]. TheScientificWorldJoumal, 2012, 2012:989514.
[78] HE B, ZHANG Y XU L, et al. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China[J]. Journal ofVirology,2014, 88(12):7070-82.
[79] PAULY M, PIR J B, LOESCH C, et al. Novel Alphacoronaviruses and Paramyxoviruses Cocirculate with Type 1 and Severe Acute Respiratory System(SARS)-Related Betacoronaviruses in Synanthropic Bats ofLuxembourg[J]. Applied and Environmental Microbiology,2017, 83(18).
[80] AR GOUILH M, PUECHMAILLE S J, DIANCOURT L, et al. SARS-CoV related Betacoronavirus and diverse Alphacoronavirus members found in western old-world[J]. Virology, 2018, 517:88-97.
[81] YANG L, WU Z, REN X, et al. Novel SARS-like betacoronaviruses in bats, China, 20 1 1[J]. Emerging Infectious Diseases, 20 13, l 9(6):989-9 1.
[82] GE X Y LI J L, YANG X L, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor[J]. Nature, 20 13, 503(7477):535-8.
[83] WANG N, LI S Y YANG X L, et al. Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China[J]. Virologica Sinica, 20 1 8, 33(1):1 04-7.
[84] WOO P C Y;LAU S K P, HUANG Yet al. Coronavirus Diversity,Phylogeny and Interspecies Jumping[J]. Experimental Biology and Medicine, 2009, 234(10):1 1 17-27.
[85] BELOUZARD S, CHU V C, WHITTAKER G R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites[J]. Proceedings of the National Academy of Sciences of the United States ofAmerica, 2009, 106(14):5871-6
[86] WU Z, YANG L, REN X, et al. ORF8-Related Genetic Evidence for Chinese Horseshoe Bats as the Source of Human Severe Acute Respiratory Syndrome Coronavirus[J]. The Journal of Infectious Diseases, 20 1 6, 2 1 3(4):579-83.
[87] HU B, ZENG L P, YANG X L, et al. Discovery of a rich gene pool of bat SARS·related coronaviruses provides new insights into the origin of SARS coronavirus[J]. PLoS Pathogens, 20 1 7,
[88] ZHOU P'LI H, WANG H, et al. Bat severe acute respiratory syndrome-like coronavirus ORF3b homologues display different interferon antagonist activities[J]. The Journal of General Virology, 2012, 93(Pt 2):275-81.
[89] YANG X L, HU B, WANG B, et al. Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus[J]. Journal ofVirology,201 5, 90(6):3253-6.
[90] MENACHERY V D, YOUNT B L, JR. , SIMS A C, et al. SARS-like WlVl·CoV poised for human emergence[J]. Proceedings of the National Academy of Sciences of the United States of America,2016, 1 13(1 1):3048-53.
[91] CSME C. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China[J]. Science(New York,NY), 2004, 303(5664):1 666·9.

[92] CHEN C Y PING Y H, LEE H C, et al. Open reading frame 8a of the human severe acute respiratory syndrome coronavirus not only promotes viral replication but also induces apoptosis[J]. The Journal of Infectious Diseases, 2007, 196(3)：405-15.
[93] KENG C T'CHOI Y Wj WELKERS M R, et al. The human severe acute respiratory syndrome coronavirus(ShRS··Coy)8b protein is distinct from its counterpart in animal SARS··CoV and down—regulates the expression of the envelope protein in infected cells[J]. Virology,2006, 354(1)： 132-42.
[94] LAW P Y LIU Y M, GENG H, et al. Expression and functional characterization of the putative protein 8b of the severe acute respiratory syndrome-associated coronavirus[J]. FEBS letters, 2006, 580(1 5)：3643-8.
[95] BALBONI A, BATTILANI M, PROSPERI S. The SAgS-like coronaviruses：the role of bats and evolutionary relationships with SARS coronavirus[J]. The New Microbiologica, 2012, 35(1)：1-16.
[96] YU P, HU B, SHI Z L, et al. Geographical structure of bat SARS-related coronaviruses[J]. Infection, Genetics and Evolution：journal of molecular epidemiology and evolutionary genetics in infectious diseases, 20 1 9, 69：224—9.
[97] ZENG L P, GAO YT, GE X Y et al. Bat Severe Acute Respiratory Syndrome—Like Coronavirus WIV l Encodes an Extra Accessory Protein, ORFX, Involved in Modulation of the Host Immune Response[J]. Journal ofVirology,201 6, 90(14)：6573—82.
[98] SOLAR／S, J. BAKER R. Mammal species of the world：a taxonomic and geographic reference [M] . 2007.
[99] CALISHER C H, CHILDS J E, FIELD H E, et al. Bats：important reservoir hosts of emerging viruses[J]. Clinical Microbiology Reviews, 2006, 1 9(3)：53 1—45
[100] SHI Z. Emerging infectious diseases associated with bat viruses[J]. Science China Life Sciences, 20 1 3, 56(8)：678—82.
[101] MORATELLI R, CALISHER C H. Bats and zoonotic viruses：can we confidently link bats with emerging deadly viruses?[J]. Memorias do Instituto Oswaldo Cruz, 2015, 110(1)：1-22.
[102] SMITH I, WANG L E Bats and their virome：an important source of emerging viruses capable ofinfecting humans[J]. Current Opinion in Virology,2013, 3(1)：84—91.
[103] CHANDY S, MATHAI D. Globally emerging hantaviruses：An overview[J1. Indian Journal ofMedical Microbiology,2017, 35(2)：165—75
[104] FISHER C R STREICKER D G, SCHNELL MJ. The spread and evolution of rabies virus： conquering new frontiers[J] -Nature Reviews Microbiology,2018, 16(4)：241—55.
[105] LAWRENCE E DANET N, REYNARD O, et al. Human transmission of Ebola virus[J] Current Opinion in Virology,20 17, 22：5 1-8
[106] CHEN T, KA—KIT LEUNG R, UU R, et al. Risk of imported Ebola virus disease in China [J]. Travel Medicine and Infectious Disease, 2014, 12(6, PartA)：650-8
[107] AWAH P K, BOOCK A U, KUM K A. Ebola Virus Diseases in Africa：a commentary on its history,local and global context[J]. The Pan African Medical Journal, 201 5, 22 Suppl l(Suppl 1)：1 8
[108] BEAN A G D, BAKER M L, STEWART C R, et al. Studying immunity to zoonotic diseases in the natural host—keeping it real[J]. Nature Reviews Immunology,20 1 3, 1 3：85 1.
[109] LUIS A D, HAyMAN D T, O'SHEA T J, et al. A comparison of bats and rodents as reservoirs of zoonotic viruses：are bats special?[J]. Proceedings Biological Sciences, 20 1 3, 280(1 756) 20122753

[110] ZHANG G, COWLED C, SHI Z, et al. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity[J]. Science(New York, NY), 20 1 3, 339(6 1 1 8) 456—60.
[111] OLIVAL K J, HOSSEINI P R, ZAMBRANA—TORRELIO C, et al. Host and viral traits predict zoonotic spillover from mammals[J]. Nature, 20 1 7, 546(7660)：646—50
[112] KATOH K, STANDLEY D M. MAFFT multiple sequence alignment software version 7 improvements in performance and usability[J]. Molecular Biology and Evolution, 20 1 3, 30(4)： 772—80.
[113] KUMAR S, STECHER G, TAMURA K. MEGA7：Molecular Evolutionary Genetics Analysis Version 7. 0 for Bigger Datasets[J]. Molecular Biology and Evolution, 2016, 33(7)：1870—4.
[114] XIA X. DAMBE5：a comprehensive software package for data analysis in molecular biology and evolution[J]. Molecular Biology and Evolution, 20 1 3, 30(7)：1 20—8
[115] MARTIN D P'MURRELL B, GOLDEN M, et al. RDP4：Detection and analysis of recombination patterns in virus genomes[J]. Virus Evolution, 2015, 1(1)：vev003一vev
[116] HUSON D H, BRYANT D. Application of phylogenetic networks in evolutionary studies [J]. Molecular Biology and Evolution, 2006, 23(2)：254-67
[117] ROBERT X, GOUET P. Deciphering key features in protein structures with the new ENDscript server[J]. Nucleic Acids Research, 20 1 4, 42(Web Server issue)：W320-4.
[118] DARRIBA D, TABOADA G L, DOALLO R, et al. jModelTest 2：more models, new heuristics and parallel computing[J]. Nature Methods, 20 12, 9(8)：772
[119] GUINDON S, DUFAYARD J F, LEFORT V et al. New algorithms and methods to estimate maximum—likelihood phylogenies：assessing the performance of PhyML 3. 0[J]. Systematic Biology, 2010, 59(3)：307—21.
[120] CONOW C, FIELDER D, OVADIA Y, et al. Jane：a new tool for the cophylogeny reconstruction problem[J]. Algorithms for Molecular Biology,2010, 5(1)：16
[121] CHARLESTON M A, ROBERTSON D L. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny[J]. Systematic Biology,2002, 5 l(3)：528—35
[122] DE SOUZA LUNA L K, HEISER V REGAMEY N, et al. Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray[J]. Journal of Clinical Microbiology,2007, 45(3)：1 049-52.
[123] CUI J, HAN N, STREICKER D, et al. Evolutionary relationships between bat coronaviruses and their hosts[J]. Emerging Infectious Diseases, 2007, 1 3(1 0)：1 526-32.
[124] MAYER F, VON HELVERSEN 0. Cryptic diversity in European bats[J]. Proceedings Biological Sciences, 2001, 268(1478)：1 825-32.

Acknowledgement

Looking back on my three years as a graduate student, I have indeed grown a lot compared to when I first came to the Virus Institute, and I have many people and things to thank.
First of all, I would like to thank my supervisors, Ms. Jie Cui and Ms. Zhengli Shi, for their patient guidance and assistance. Thank you for your guidance on my topic and academic ideas. Your love and rigorous attitude towards scientific research have greatly influenced me and benefited me a lot.
I would like to thank Ms. Hu Yi for her guidance and help in my experiments, result analysis and English writing. I thank Mr. Libiao Zhang, Mr. Xingwen Peng, Mr. Yi Fan, Mr. Dongsheng Luo, and Yun Luo for their hard sampling work in the early stage. We thank Ms. Li Bei, Ms. Zhang Wei, Ms. Zhu Yan, Ms. Chen Jing, Ms. Wang Ning, and Zheng Xiaoshuang for their guidance and help in experimental operation. I would like to thank Mr. Lei Zhang from the Analysis and Testing Center for his explanation and demonstration of the high throughput library building process, which gave me an additional understanding of the front end of high throughput analysis. I would like to thank Ms. Chen Mingyue, Ms. Fang Wei and Ms. Li Chunxuan for their help and care in life and study; and friends such as Liu Lu and Wang Wenfang for their mutual support and mutual encouragement in life and study.
Thanks to some like-minded partners I have met in these three years. Thanks to Zhang Guangyang, a postdoctoral fellow from the University of Florida, Shen Xingxing, a postdoctoral fellow from Vanderbilt University, Professor Yu Guangchuang from Southern Medical University, and LQ from Huazhong Agricultural University for sharing, helping and discussing in the study of raw materials; thanks to Gao Fangluan from Fujian Agricultural University for sharing the courseware and experience summaries as well as guidance and help when consulting problems; thanks to some useful public numbers such as biobabble I would like to thank some useful public websites such as biobabble, biobabble skill tree, biobabble daughter, science gaga club, etc. for sharing their learning posts, and Github, Biostars and other learning websites for their help in daily learning.
I would like to thank Ms. Wang Yanfei, Ms. He Man, Ms. Pei Pingping and Ms. Wang Dan for their care and concern during my three years of graduate life.
Lastly, I would like to thank my parents and family members for their support, encouragement and care behind me, so that I can freely choose and courageously do what I want to do, so that I can have no worries on my study path. At the same time, I would like to thank myself for always persisting in my studies.
The breeze is blowing, the cherry blossoms are blooming, the sun is shining, and I wish you all the best for the future!

Yu Ping
April 2019

Author's Biography and Academic Papers and Research Results Published during the Degree Pursuit

Author's Biography:
She received her B.S. degree from the College of Life Sciences, Hubei University, September 2012 one by one, June 2016.
She received her M.S. degree from Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China, September 2016, June 2019.

Published (or formally accepted) academic papers:
1. **Yu P**, Hu B, Shi ZL, Cui J. 20 1 9. Geographical structure of bat SARS-related Geographical structure of bat SARS-related coronaviruses. Infect Genet Evol 69: 224-229. (co-author)

Awards:
In 2017, she was awarded "Outstanding Volunteer" by the Wuhan Education Base of Chinese Academy of Sciences, Xiaohongshan Youth Volunteer Association.
In 2018, she won the second prize of poster presentation in the 7th Virology Postgraduate Forum.