Homologous recombination within the spike glycoprotein of the newly identified coronavirus 2019-nCoV may boost cross-species transmission from snake to human

Wei Ji[1, *, #], Wei Wang[2, *], Xiaofang Zhao[3, *], Junjie Zai [4, *], Xingguang Li[5, *]

1. Department of Microbiology, Peking University Health Science Center School of Basic Medical Sciences, Beijing, China; jiwei_yunlong@126.com
2. Department of Spleen and Stomach Diseases, The First affiliated Hospital of Guangxi university of Chinese Medicine, Nanning 530023, China.
3. Department of Science and Technology, Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning 530011, China
4. Immunology innovation team, School of Medicine, Ningbo University, Ningbo 315211, China.
5. Hubei Engineering Research Center of Viral Vector, Wuhan University of Bioengineering, Wuhan 430415, China.

*These authors contributed equally to this work.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/fut.22099.

Keywords

2019-nCoV; coronavirus; phylogenetic analysis; recombination; spike glycoprotein; codon usage bias; cross-species transmission; snake

Abstract

The current outbreak of viral pneumonia in the city of Wuhan, China, was caused by a novel coronavirus designated 2019-nCoV by the World Health Organization, as determined by sequencing the viral RNA genome. Many initial patients were exposed to wildlife animals at the Huanan seafood wholesale market, where poultry, snake, bats, and other farm animals were also sold. To determine possible virus reservoir, we have carried out comprehensive sequence analysis and comparison in conjunction with relative synonymous codon usage (RSCU) bias among different animal species based on the 2019-nCoV sequence. Results obtained from our analyses suggest that the 2019-nCoV appears to be a recombinant virus between the bat coronavirus and an origin-unknown coronavirus. The recombination occurred within the viral spike glycoprotein, which recognizes cell surface receptor. Additionally, our findings suggest that snake is the most probable wildlife animal reservoir for the 2019-nCoV based on its RSCU bias close to snake compared to other animals. Taken together, our results suggest that homologous recombination within the spike glycoprotein may contribute to the 2019-nCoV cross-species transmission from snake to humans.

Introduction

China has been the epicenter of emerging and reemerging viral infections that continue to stir a global concern. In the last 20 years, China has witnessed several emerging viral diseases, including an avian influenza in 1997[1], the severe acute respiratory syndrome (SARS) in 2003[2], and a severe fever with thrombocytopenia syndrome (SFTS) in 2010[3]. The most recent crisis was the outbreak of an ongoing viral pneumonia with unknown etiology in the city of Wuhan, China. On December 12, 2019, Wuhan Municipal Health Commission (WMHC) reported 27 cases of viral pneumonia with 7 of them being critically ill. Most of them had history of exposure to the virus at the Huanan Seafood Wholesale Market where poultry, bats, snakes; and other wildlife animals were also sold[4]. On January 3rd, 2020, WMHC updated the number of cases to a total of 44 with 11 of them in critical condition. On January 5th, the number of cases increased to 59 with 7 critically ill patients. The viral pneumonia outbreak was not caused by severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East Respiratory Syndrome coronavirus (MERS-CoV), influenza virus, or adenovirus as determined by laboratory tests[4]. On January 10, it was reported that a novel coronavirus designated 2019-nCoV by the World Health Organization (WHO)[5] was identified by high-throughput sequencing of the viral RNA genome, which was released through virological.org. More significantly, the newly identified 2019-CoV has also been isolated from one patient. The availability of viral RNA sequence has made it possible to develop reverse-transcription polymerase chain reaction (RT-PCR) methods for detection of viral RNA in samples from patients and potential hosts[6]. As a result, 217 patients were confirmed to be infected with the 2019-nCoV, and 9 patients died as of January 20, 2020. Several patients from Wuhan were also reported in Thailand, Singapore, Hong Kong, South Korea and Japan. High-thrpughput sequencing of viral RNA from patients' samples has identified a novel coronavirus designated 2019-nCoV by the World Health Organization.

Currently, a total of 14 full-length sequences of the 2019-nCoV were released to GISAID and GeneBank.

The *coronavirinae* family consists of four genera based on their genetic properties, including genus *Alphacoronavirus*, genus *Betacoronavirus*, genus *Gammacoronavirus*, and genus *Deltacoronavirus*[7]. The coronavirus RNA genome (ranging from 26 to 32 kb) is the largest among all RNA viruses[8]. Coronavirus can infect humans and many different animal species, including swine, cattle, horses, camels, cats, dogs, rodents, birds, bats, rabbits, ferrets, mink, snake, and other wildlife animals[7,9]. Many coronavirus infections are subclinical[7,9]. The severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus(MERS-CoV) belong to the *Betacoronavirus* genus and are zoonotic pathogens that can cause severe respiratory diseases in humans[7].

The outbreak of viral pneumonia in Wuhan is associated with history of exposure to virus reservior at the Huanan seafood wholesale market, suggesting a possible zoonosis. The seafood market also sold live animals such as snakes, marmots, bats, birds, frogs, hedgehogs and rabbit. Currently, there is no evidence suggesting a specific wildlife host as virus reservoir. Studies of relative synonymous codon usage (RSCU) bias between viruses and their hosts suggested that viruses tends to evolve codon usage bias that are comparable to their hosts[10,11]. To search for potential virus reservoir, we have carried out a comprehensive sequence analysis and comparison. Results from our analysis suggest that snake is the most probable wildlife animal reservoir responsible for the current outbreak of 2019-nCoV infection. More interestingly, an origin-unknown homologous recombination was identified within the spike glycoprotein of the 2019-nCoV[5], which may explain its decreased pathogenesis, snake-to-human cross species transmission, and limited person-person spread.

Materials and Methods

Sequence data collection

The newly sequenced Beta-coronavirus (MN908947) genome was downloaded from the GenBank database. Five hundred closely related sequences were also downloaded from GenBank. Out of them, 271 genome sequences (>19,000bp in length) were used in this study together with the above-described Beta-coronavirus (2019-nCoV, MN908947) genome sequence (Supplementary Table S1). The geographic origins of the sequences were from Bulgaria (*n*=1), Canada (*n*=2), China (n=67), Germany (n=1), Hong Kong (n=5), Italy (n=1), Kenya (n=1), Russia (n=1), Singapore (n=24), South Korea (n=1), Taiwan (n=11), United Kingdom (n=2), United States of America (*n*=67), and Unknown (n=88). Sequences were aligned using MAFFT v7.222[12], followed by manual adjustment using BioEdit v7.2.5[13].

Phylogenetic and recombination analysis

Phylogenetic trees were constructed using maximum-likelihood methods and general time-reversible model of nucleotide substitution with gamma-distributed rates among sites (GTR+G substitution model) in RAxML v8.0.9[14]. Support for the inferred relationships was evaluated by a bootstrap analysis with 1000 replicates and trees were midpoint-rooted.

To investigate the putative parents of the 2019-nCoV, we performed Similarity and Bootscanning plot analyses based on the Kimura two-parameter model with window size of 500 bp, step size of 30 bp using SimPlot v.3.5.1[15]. We divided our data set into 4 clades, the newly discovered 2019-nCoV sequence were grouped as the query sequence. The closest relative coronaviruses (bat-SL-CoVZC45 and bat-SL-CoVZXC21) obtained from the city of Nanjing, China were grouped as ‘Clade A’. The other two coronaviruses (BtCoV/BM48-31/BGR/2008 and BtKY72) from

Bulgaria and Kenya were grouped as ‘Clade B’. The rest sequences were grouped as ‘Clade C’ (Fig. 1).

Synonymous Codon Usage Analysis

To estimate the relative synonymous codon usage (RSCU) bias of the 2019-nCoV and its potential host(s), coding sequences of the 2019-nCoV genome (9672 codons), bat-SL-CoVZC45 genome (9680 codons), Bungarus multicinctus genes (32789 codons), Naja Atra genes (30270 codons), Erinaceus europaeus genes (16967352 codons), Marmota genes (21276356 codons), manis javanica genes (24183590 codons), Rhinolophus sinicus genes (28430 codons) and Gallus gallus genes (314483 codons) from GenBank were calculated with Codon W1.4.2 [16,17]. The RSCU of human genes (40662582 codons) was retrieved from the Codon Usage Database (http://www.kazusa.or.jp/codon/). The relationship among these sequences was calculated using a squared euclidean distance ($d_{ik} = \sum_{j=1}^{p}(X_{ij} - X_{kj})^2$), as we previously reported[18]. A heat map of RSCU was drawn with MeV 4.9.0 software[19]. The Coronavirus and their potential hosts were clustered using a Euclidean distance method.

Results

Phylogenetic classification

Phylogenetic analysis of 276 coronavirus genomes revealed that the newly identified coronavirus 2019-nCoV sequence was monophyletic with 100% bootstrap support. The Clade A (bat-SL-CoVZC45 and bat-SL-CoVZXC21) derived from bats in the city of Nanjing, China between 2015-2017 represents the sister lineage to 2019-nCoV. The Clade B (BtCoV/BM48-31/BGR/2008 and BtKY72) obtained from bats in Bulgaria and Kenya between 2005-2007 formed a distinct monophyletic cluster with 100% bootstrap support. The Clade C including 267 coronavirus strains was clustered together with 63% bootstrap support (Fig. 1).

Homologous recombination within the viral spike glycoprotein may boost Cross-species transmission.

Homologous recombination is an important evolutionary force and previous studies have found that homologous recombination occurred in many viruses, including Dengue virus[20], human immunodeficiency virus[21], hepatitis B virus[22], hepatitis C virus[23] and classical swine fever virus[18]. Similarity plot analysis of the 2019-nCoV revealed that homologous recombination occurred between Clade A strains (bat-coronaviruses) and the origin-unknown isolates in 21500-24000bp, located within the spike glycoprotein that recognizes cell surface receptor (Fig. 2). These characteristics indicates that homologous recombination within the spike glycoprotein may boost the 2019-nCoV cross-species spread to humans.

Snakes as the most probable wildlife animal reservoir for the 2019-nCoV.

As parasitic microorganism, virus codon usage pattern resembles its host to some extent. The RSCU bias shows that the 2019-nCoV, bat-SL-CoVZC45, and snakes from China have similar synonymous codon usage bias (Fig. 3A, Table1). The squared euclidean distance indicates that the 2019-nCoV and snakes from China have the highest similarity in synonymous codon usage bias compared to those of marmota, hedgehog, manis, bat, bird and human (Fig. 3B). Two types of snakes, containing Bungarus multicinctus (Many-banded krait) and Naja Atra (Chinese cobra) were used for RSCU analysis. Squared euclidean distance between the 2019-nCoV and Bungarus multicinctus is 12.47. The distance between the 2019-nCoV and another snake Naja Atra is 14.70. However, the distance between the 2019-nCoV and other animals is greater than 24, specifically 24.87 for marmota, 25.92 for hedgehog, 26.81

for manis, 27.47 for bat, 29.07 for bird, and 35.44 for human. These data suggest that the 2019-nCoV can more effectively use snake's translation machinery than that of other animals, suggesting that the epidemic 2019-nCoV might originate from snakes.

Two types of snakes are common in the Southeastern China including the city of Wuhan (Fig. 4). Geographical distributions of Bungarus multicinctus include Taiwan, the Central and Southern China, Hong Kong, Myanmar (Burma), Laos, and Northern Vietnam[24]. Naja Atra is found in Southeastern China, Hong Kong, Northern Laos, Northern Vietnam, and Taiwan[25]. Snakes were also sold at the Huanan Seafood Wholesale Market where many patients worked or had history of exposure to wildlife or farm animals. Taken together, snakes could be the most probable wildlife animal reservoir for the 2019-nCoV.

Discussion

In this study, we have performed an evolutionary analysis using 272 genomic sequences of coronaviruses obtained from various geographic locations and host species. Our results show that the novel coronavirus sequence obtained from the viral pneumonia outbreak occurring in the city of Wuhan form a separate group that is highly distinctive to SARS-CoV. The SARS-CoV first emerged in China in 2002 and then spread to 37 countries/regions in 2003 and caused a travel-related global outbreak with 9.6% mortality rate [26]. More importantly, results from our analysis reveal a homologous recombination occurred between the bat coronavirus and an origin-unknown coronavirus in the nucleotides 21500-24000 within the viral spike glycoprotein gene. Sequence homology analysis of the partial spike glycoprotein genes (1-783bp) from the 2019-nCoV was done through BLAST at the NCBI website. Interestingly, no similar sequence was found with known sequence in the database, suggesting that a putative recombination parent virus was still unknown. Previous study suggested that recombination of SARS in the spike glycoprotein genes might have mediated the initial cross-species transmission event from bats to other mammals[27]. Bootscanning plot analysis (data not shown) suggested that the major parents of the 2019-nCoV originated from Clade A (bat-SL-CoVZC45 and bat-SL-CoVZXC21) but formed a monophyletic cluster different from them. A possible explanation for the monophyly could be insertion of the spike glycoprotein regions of another parent lineage into Clade A. Overall, the ancestral origin of the 2019-nCoV was more likely from divergent host species rather than SARS-CoV.

The host range of some animal coronaviruses was promiscuous[7]. They caught our attention only when they caused human diseases such as SARS, MERS and 2019-nCoV pneumonia[4,9,28]. It is critical to determine the animal reservoir of the 2019-nCoV in order to understand the molecular mechanism of its cross-species spread. Homologous recombination within viral structural proteins between coronaviruses from different hosts may be responsible for “cross-species” transmission[27]. Information obtained from RSCU analysis provides some insights to the question of wildlife animal reservoir although it requires further validation by experimental studies in animal models. Currently, the 2019-nCoV has not been isolated from animal species although it was obtained from one patient. Identifying and characterizing the animal reservoir for 2019-nCoV will lead to the determination of the recombination parent sources and for a better understanding of its person-to-person spread among human populations.

The 2019-nCoV has caused a total of 217 confirmed cases of pneumonia in China as of January 20, 2020 with new patients also reported in Hong Kong, Thailand, Singapore, South Korea and Japan. Unlike SARS-CoV, the 2019-nCoV appeared to initially cause mild form of viral pneumonia and have limited capability for person-person spread. This could be due to the recombination occurred within the receptor-binding glycoprotein, virus origination from snake, or both. Snakes are cold-blooded reptiles with lower temperature than humans[29]. Accordingly, the 2019-nCoV will likely be attenuated upon infection to humans. However, there is a concern about its adaptation in humans that may acquire the capability to replicate more efficiently and spread more rapidly via close person-person contact.

In summary, results derived from our evolutionary analysis suggest for the first time that snake is the most probable wildlife animal reservoir for the 2019-nCoV based on their similar codon usage bias. Additionally, our analyses identified a homologous recombination within the viral receptor-binding spike glycoprotein, which may determine cross-species transmission from snake to humans. These novel findings warrant future investigation to experimentally determine if snake serves as the 2019-nCoV reservoir and the homologous recombination within the spike glycoprotein determine the tropism of the 2019-nCoV in viral transmission and replication. New information obtained from our evolutionary analysis is highly significant for effective control of the outbreak caused by the 2019-nCoV-induced pneumonia.

References

1 Chan, P. K. Outbreak of avian influenza A (H5N1) virus infection in Hong Kong in 1997. *Clinical Infectious Diseases* **34**, S58-S64 (2002).

2 Guan, Y. *et al.* Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276-278 (2003).

3 Yu, X. J. *et al.* Fever with thrombocytopenia associated with a novel bunyavirus in China. *N Engl J Med* **364**, 1523-1532, doi:10.1056/NEJMoa1010095 (2011).

4 Lu, H., Stratton, C. W. & Tang, Y. W. Outbreak of Pneumonia of Unknown Etiology in Wuhan China: the Mystery and the Miracle. *J Med Virol*, doi:10.1002/jmv.25678 (2020).

5 WHO. *Coronavirus*, <https://www.who.int/health-topics/coronavirus> (2019).

6 Zhang, N. *et al.* Recent advances in the detection of respiratory virus infection in humans. *J*

Med Virol, doi:10.1002/jmv.25674 (2020).

7 MacLachlan, N. J. & Dubovi, E. J. in *Fenner's Veterinary Virology (Fifth Edition)* (eds N. James MacLachlan & Edward J. Dubovi) 393-413 (Academic Press, 2017).

8 Howley, D. M. K. P. *Fields virology*. Vol. 1 830P (Lippincott Williams & Wilkins (LWW), 2013).

9 Organization, W. H. Consensus document on the epidemiology of severe acute respiratory syndrome (SARS). (Geneva: World Health Organization, 2003).

10 Wang, H., Liu, S., Zhang, B. & Wei, W. Analysis of synonymous codon usage bias of Zika virus and its adaption to the hosts. *PLoS One* **11**, e0166260 (2016).

11 Bahir, I., Fromer, M., Prat, Y. & Linial, M. Viral adaptation to host: a proteome‐based analysis of codon usage and amino acid preferences. *Molecular systems biology* **5** (2009).

12 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

13 Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98, doi:citeulike-article-id:691774 (1999).

14 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

15 Lole, K. S. *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152-160 (1999).

16 Liu, X., Zhang, Y., Fang, Y. & Wang, Y. Patterns and influencing factor of synonymous codon usage in porcine circovirus. *Virology journal* **9**, 1-9 (2012).

17 Wright, F. The ‘effective number of codons’ used in a gene. *Gene* **87**, 23-29 (1990).

18 Ji, W., Niu, D.-D., Si, H.-L., Ding, N.-Z. & He, C.-Q. Vaccination influences the evolution of classical swine fever virus. *Infection, Genetics and Evolution* **25**, 69-77 (2014).

19 Howe, E. *et al.* in *Biomedical Informatics for Cancer Research* (ed Michael F. OchsJohn T. CasagrandeRamana V. Davuluri) 267-277 (2010).

20 Tolou, H. *et al.* Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. *Journal of General Virology* **82**,

1283-1290 (2001).

21 Clavel, F. *et al.* Genetic recombination of human immunodeficiency virus. *Journal of virology* **63**, 1455-1459 (1989).

22 Bollyky, P. L., Rambaut, A., Harvey, P. H. & Holmes, E. C. Recombination between sequences of hepatitis B virus from different genotypes. *Journal of Molecular Evolution* **42**, 97-102 (1996).

23 Colina, R. *et al.* Evidence of intratypic recombination in natural populations of hepatitis C virus. *Journal of general virology* **85**, 31-37 (2004).

24 wikipedia. *Many-banded krait*, <https://en.wikipedia.org/wiki/Many-banded_krait> (2020).

25 wikipedia. *Chinese cobra*, <https://en.wikipedia.org/wiki/Chinese_cobra> (2020).

26 Smith, R. D. Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. *Social science & medicine* **63**, 3113-3123 (2006).

27 Graham, R. L. & Baric, R. S. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *Journal of virology* **84**, 3134-3146 (2010).

28 Cunha, C. B. & Opal, S. M. Middle East respiratory syndrome (MERS) A new zoonotic viral pneumonia. *Virulence* **5**, 650-654 (2014).

29 Hayes, R. O., Daniels, J. B., Maxfield, H. K. & Wheeler, R. E. Field and laboratory studies on eastern encephalitis in warm-and cold-blooded vertebrates. *The American journal of tropical medicine and hygiene* **13**, 595-606 (1964).

Acknowledgements

This work was supported by Project of Guangxi Health Committee (No. Z20191111) and Natural Science Foundation of Guangxi Province of China (No. 2017GXNSFAA198080) to Dr. Xiaofang Zhao. This study was sponsored by K.C. Wong Magna Fund in Ningbo University.

We want to thank Prof. Yongzhen Zhang (Fudan University) for deposing the sequence of the newly identified coronavirus 2019-nCoV to GeneBank, which was used in this study.

Author contributions

Conceptualization: Wei Ji

Writing: Wei Ji, Xingguang Li

Data collection: Junjie Zai, Wei Wang, Xiaofang Zhao

Data analysis: Wei Ji, Xingguang Li

Competing financial interests

The authors declare no competing financial interests.

Figure legends

Figure 1. Maximum likelihood phylogenetic tree of the 2019-nCoV.

Phylogenetic tree inferred from 272 near-complete genome sequences of coronavirus was midpoint rooted and grouped into 4 clades (2019-nCoV, Clades A, B, and C). Coronaviruses originating from different countries/regions are highlighted in colors.

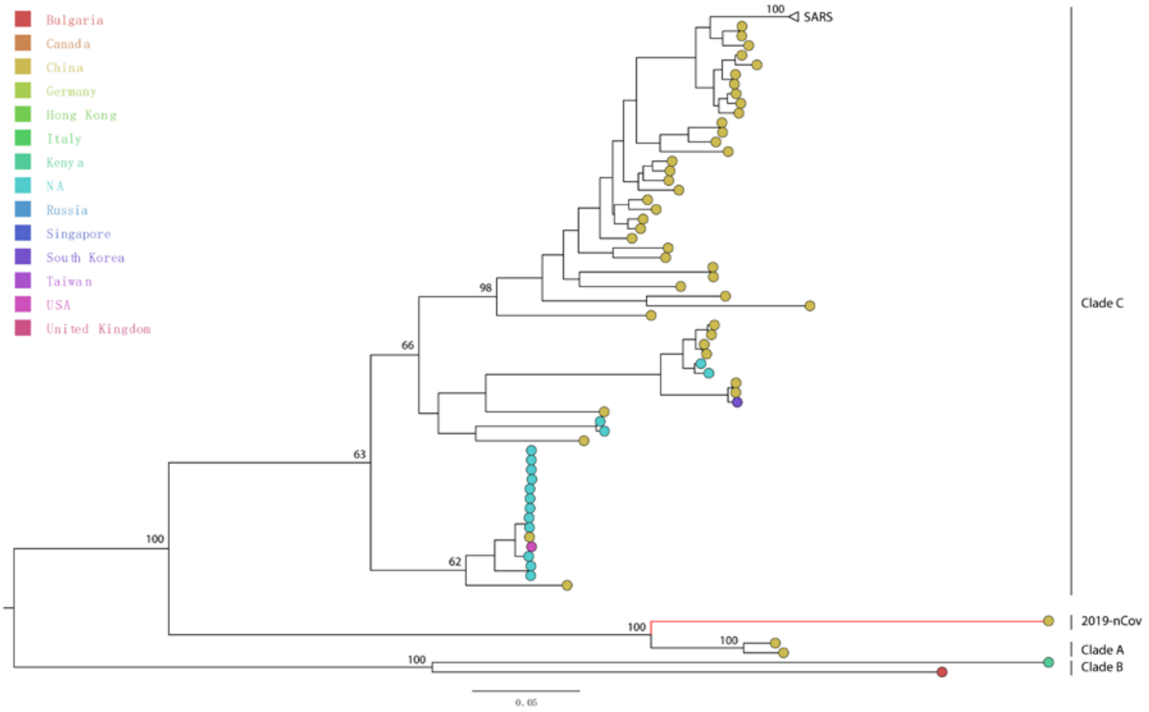

Figure 2. Sequence comparison among different coronaviruses. Similarity plot analysis was performed among coronaviruses in Clades A, B, and C. Recombination analysis was conducted with a sliding window of 500 bp and a step size of 30 bp. Recombination sites were located within the viral spike glycoprotein genes, as indicated by an orange box on the top.

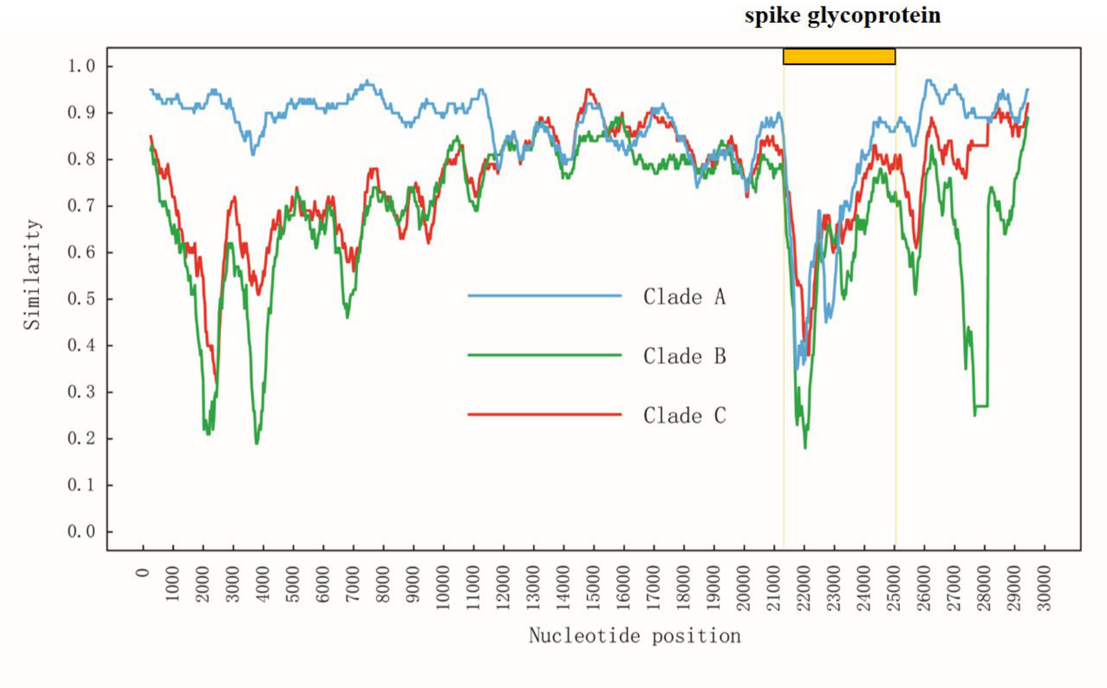

Figure 3 Comparison of relative synonymous codon usage (RSCU) between 2019-nCoV and its putative wildlife animal reservoir(s). A. Heat map resulting from cluster analysis of the RSCU among the 2019-nCoV, bat-SL-CoVZC45, Bungarus multicinctus, Naja atra, Marmota, Erinaceus europaeus, manis javanica, Rhinolophus sinicus, Gallus gallus, Homo sapiens. **B**. Comparison of squared euclidean distance between 2019-nCoV and different animal species. Squared euclidean distance was calculated based on the RSCU.

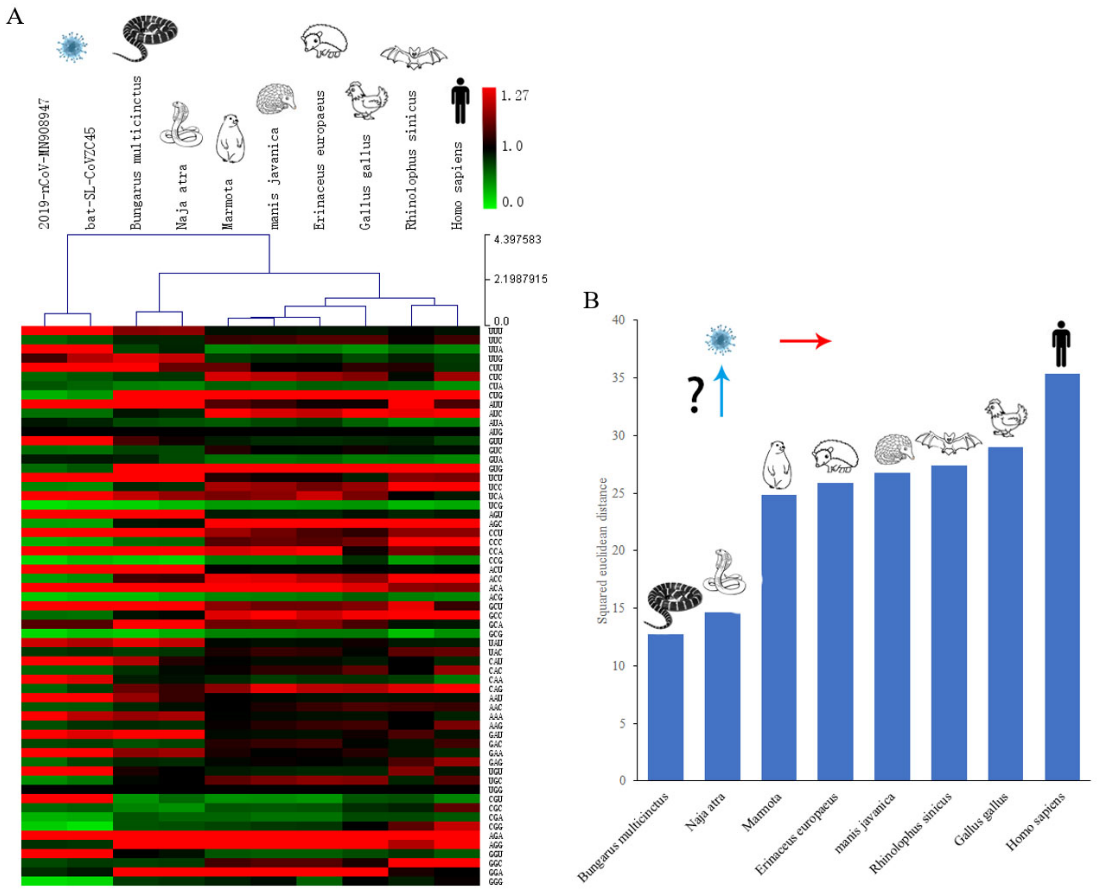

Figure 4. Geographic distribution of Bungarus multicinctus and Naja atra in China. The geographic distribution of Bungarus multicinctus and Naja atra are highlighted in colors. Yellow color represents the common geographic distribution of Bungarus multicinctus and Naja atra. Green color represents additional geographic distribution of Bungarus multicinctus. The location of Wuhan city where the 2019-nCoV outbreak occurs is indicated in red. Maps were obtained from Craft MAP website (http://www.craftmap.box-i.net/).

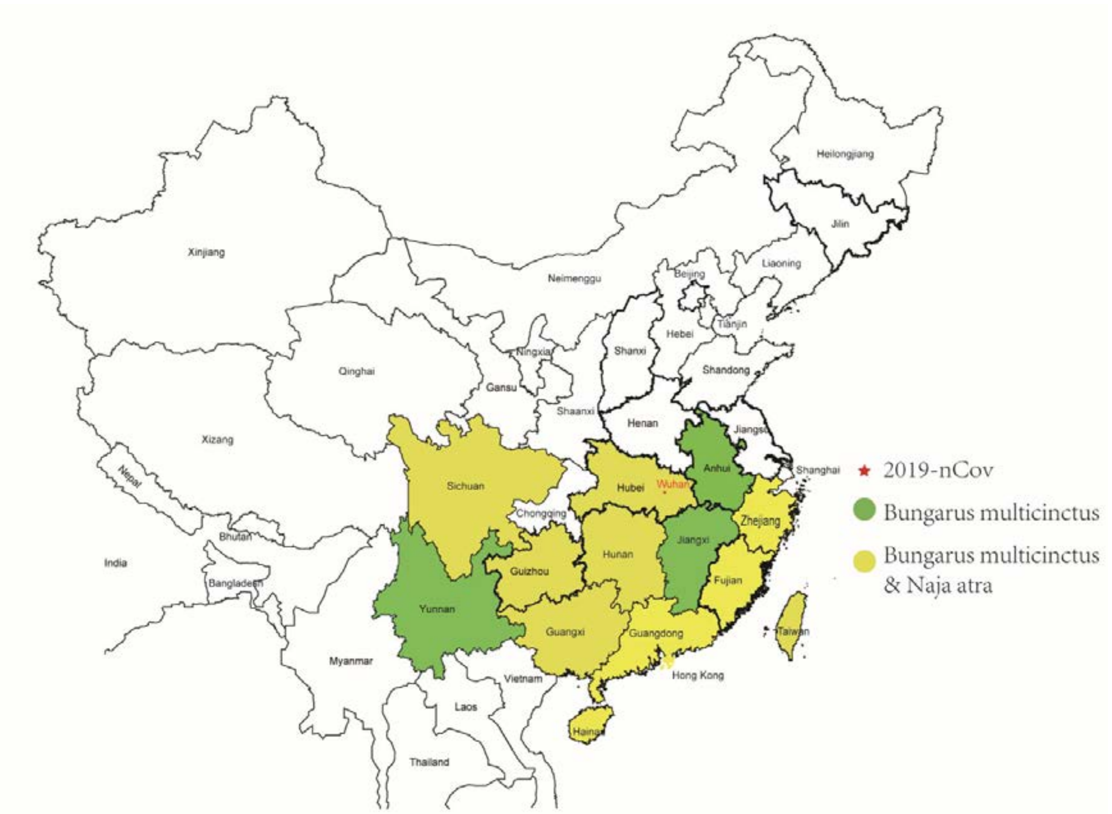

Table1 The RSCU analysis of the preferred codons (codons with RSCU > 1), the optimal codons and the rare codons for coronoviruses, snakes, hedgehog, bat Marmota, Manis, gallus, and human genome. The most preferred codons are in bold or red.

		bat-SL-CoVZC45	2019-nCoV-MN908947	Bungarus multicinctus	Naja atra	Marmota	Erinaceus europaeus	manis javanica	Rhinolophus sinicus	Gallus gallus	Homo sapiens
Phe	UUU	**1.33**	**1.41**	**1.14**	**1.15**	0.94	0.91	0.93	0.99	0.91	0.93
	UUC	0.67	0.59	0.86	0.85	**1.06**	**1.09**	**1.07**	**1.01**	**1.09**	**1.07**
Leu	UUA	1.37	1.64	0.73	0.86	0.46	0.49	0.48	0.56	0.42	0.46

	UUG	1.19	1.07	1.24	1.21	0.77	0.88	0.85	0.87	0.72	0.77
	CUU	**1.77**	**1.75**	1.3	1.11	1.11	1.00	1	1.04	1.05	0.79
	CUC	0.66	0.59	0.78	0.79	1.23	1.17	1.2	0.96	1.14	1.17
	CUA	0.6	0.66	0.5	0.47	0.62	0.63	0.57	0.57	0.6	0.43
	CUG	0.4	0.3	**1.45**	**1.55**	**1.81**	**1.83**	**1.89**	**2**	**2.06**	**2.37**
II	A	**1.57**	**1.53**	**1.37**	**1.**	1.09	1.02	1.0	1.28	1.0	1.0

e	UU				**46**			6		2	8
	AUC	0.56	0.56	0.92	0.86	**1.27**	**1.21**	**1.23**	**1.25**	**1.36**	**1.41**
	AUA	0.86	0.91	0.71	0.68	0.64	0.77	0.71	0.47	0.61	0.51
Met	AUG	1	1	1	1	1	1.00	1	1	1	1.00
Val	GUU	**1.89**	**1.95**	1.08	1.02	0.89	0.82	0.83	0.87	0.85	0.73
	GU	0.55	0.57	0.74	0.7	1.04	**0.99**	1	**0.95**	1.02	**0.95**

	C				1						
	GUA	0.91	0.9	0.88	0.71	0.56	0.59	0.56	0.53	0.51	0.47
	GUG	0.66	0.58	**1.29**	**1.56**	**1.51**	**1.60**	**1.61**	**1.66**	**1.62**	**1.85**
Ser	UCU	**2.04**	**1.96**	**1.5**	**1.43**	1.04	1.02	1.02	1.14	0.88	1.13
	UCC	0.44	0.47	0.85	0.81	1.18	1.12	1.16	1.31	1.16	1.31
	UCA	1.66	1.66	1.21	1.16	1.13	1.22	1.15	1	1.13	0.90

	UCG	0.15	0.11	0.17	0.26	0.39	0.40	0.38	0.28	0.41	0.33
	AGU	1.36	1.43	1.35	1.39	0.89	0.86	0.87	0.85	0.93	0.90
	AGC	0.36	0.37	0.91	0.93	**1.37**	**1.38**	**1.41**	**1.42**	**1.5**	**1.44**
Pro	CCU	**1.82**	**1.94**	**1.7**	**1.66**	1.17	1.08	1.12	1.15	1.05	1.15
	CCC	0.34	0.3	0.57	0.59	1.09	1.09	1.11	**1.35**	**1.12**	**1.29**
	C	1.59	1.6	1.51	1.	**1.23**	**1.27**	**1.2**	1.13	1.0	1.1

	CA				58			**5**		2	1
	CCG	0.26	0.16	0.21	0.17	0.51	0.56	0.51	0.37	0.81	0.45
Thr	ACU	**1.75**	**1.78**	1.3	1.28	1	0.97	0.97	0.96	0.99	0.99
	ACC	0.44	0.38	1.07	1.06	1.25	1.20	1.24	**1.39**	1.15	**1.42**
	ACA	1.58	1.64	**1.38**	**1.33**	**1.29**	**1.31**	**1.31**	1.15	**1.24**	1.14
	AC	0.24	0.2	0.24	0.3	0.47	0.52	0.49	0.5	0.62	0.46

	G				4						
Ala	GCU	**2.13**	**2.19**	**1.54**	1.34	1.15	1.12	1.12	1.25	1.14	1.06
	GCC	0.55	0.57	0.93	0.97	**1.24**	**1.22**	**1.27**	**1.59**	**1.26**	**1.60**
	GCA	1.09	1.09	1.33	**1.38**	1.12	1.14	1.13	0.93	1.08	0.91
	GCG	0.24	0.15	0.2	0.31	0.48	0.52	0.48	0.24	0.52	0.42
Tyr	UAU	**1.19**	**1.22**	**1.25**	**1.21**	0.97	0.95	0.94	0.9	**1.02**	0.89

	UAC	0.81	0.78	0.75	0.79	**1.03**	**1.05**	**1.06**	**1.1**	0.98	**1.11**
His	CAU	**1.39**	**1.39**	**1.19**	**1.04**	0.99	0.95	0.95	1	0.9	0.84
	CAC	0.61	0.61	0.81	0.96	**1.01**	**1.05**	**1.05**	1	**1.1**	**1.16**
Gln	CAA	**1.24**	**1.39**	0.89	0.94	0.85	0.80	0.75	0.76	0.81	0.53
	CAG	0.76	0.61	**1.11**	**1.06**	**1.15**	**1.20**	**1.25**	**1.24**	**1.19**	**1.47**
A	A	**1.34**	**1.35**	**1.16**	**1.**	1	0.94	0.9	0.93	0.9	0.9

sn	AU				**06**			7		2	4
	AAC	0.66	0.65	0.84	0.94	1	**1.06**	**1.03**	**1.07**	**1.08**	**1.06**
Lys	AAA	**1.2**	**1.31**	**1.16**	**1.18**	0.99	0.94	0.96	1	0.93	0.87
	AAG	0.8	0.69	0.84	0.82	**1.01**	**1.06**	**1.04**	1	**1.07**	**1.13**
Asp	GAU	**1.24**	**1.28**	**1.32**	**1.27**	0.96	0.93	0.94	**1.08**	**1.02**	0.93
	GA	0.76	0.72	0.68	0.7	**1.04**	**1.07**	**1.06**	0.92	0.98	**1.07**

	C				3						
Glu	GAA	**1.27**	**1.44**	**1.15**	**1.17**	**1.03**	**1.01**	0.99	0.92	**1.04**	0.84
	GAG	0.73	0.56	0.85	0.83	0.97	0.99	**1.01**	**1.08**	0.96	**1.16**
Cys	UGU	**1.47**	**1.56**	**1.03**	1	0.9	0.85	0.87	**1.14**	0.86	0.91
	UGC	0.53	0.44	0.97	1	**1.1**	**1.15**	**1.13**	0.86	**1.14**	**1.09**
Trp	UGG	1	1	1	1	1	1.00	1	1	1	1.00

Arg	CGU	1.52	1.45	0.42	0.59	0.43	0.41	0.44	0.69	0.66	0.48
	CGC	0.63	0.59	0.48	0.43	0.66	0.67	0.65	0.87	0.82	1.10
	CGA	0.32	0.29	0.66	0.6	0.6	0.55	0.55	0.72	0.71	0.65
	CGG	0.1	0.19	0.67	0.69	0.73	0.79	0.78	1.1	0.99	1.21
	AGA	**2.63**	**2.67**	**2.29**	**2.07**	**2.07**	**1.97**	**1.94**	**1.42**	**1.46**	**1.29**
	A	0.79	0.81	1.47	1.	1.51	1.61	1.6	1.19	1.3	1.2

	GG				61			3		6	7
Gly	GGU	**2.16**	**2.34**	0.99	0.93	0.66	0.63	0.63	0.79	0.7	0.65
	GGC	0.77	0.71	0.75	0.77	1.07	1.09	1.09	**1.39**	1.03	**1.35**
	GGA	0.91	0.83	**1.59**	**1.64**	**1.39**	**1.33**	**1.31**	1.03	**1.29**	1.00
	GGG	0.16	0.12	0.77	0.81	0.88	0.62	0.95	0.88	0.99	1.01