

# An Update on the Origin of SARS-CoV-2: Though Closest Identity, Bat (RaTG13) and Pangolin Derived Coronaviruses Varied in the Critical Binding Site and O-Linked Glycan Residues

Malaiyan Jeevan (✉ [jeevan1209@gmail.com](mailto:jeevan1209@gmail.com))

Sri Muthukumaran Medical College Hospital and Research Institute <https://orcid.org/0000-0001-6466-308X>

Arumugam Suresh

Meenakshi Medical College Hospital and Research Institute

Mohan Kamalraj

Sri Muthukumaran Medical College Hospital and Research Institute

Gomathi Radhakrishnan Gokul

Sri Muthukumaran Medical College Hospital and Research Institute

---

## Research

**Keywords:** COVID-19, intermediate host, pangolins, RaTg13, SARS-CoV-2

**DOI:** <https://doi.org/10.21203/rs.3.rs-28534/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** The initial cases of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) occurred in Wuhan, China, in December 2019 and swept the world by 13 May 2020 with 4,179,479 active cases, 287,525 deaths across 215 countries, areas or territories. This strongly implies global transmission occurred before the lockdown of China. However, the initial source's transmission routes of SARS-CoV-2 remain obscure and controversial. Research data suggest bat (RaTG13) and pangolin carried CoV were the proximal source of SARS-CoV-2.

**Methods** In this study, we used systematic phylogenetic analysis of *Coronavirinae* subfamily along with wild type human SARS-CoV, MERS-CoV, and SARS-CoV-2 strains. The key residues of the receptor-binding domain (RBD), O-linked glycan and Angiotensin-converting enzyme 2 (ACE2) were compared.

**Results** SARS-CoV-2 strains were clustered with RaTG13 (97.41% identity), Pangolin-CoV (92.22% identity) and Bat-SL-CoV (80.36% identity), forms a new clade-2 in lineage B of beta-CoV. The alignments of RBD contact residues to ACE2 justified those SARS-CoV-2 strains sequences were 100% identical by each other, significantly varied in RaTG13 and pangolin-CoV. SARS-CoV-2 has a polybasic cleavage site with an inserted sequence of PRRRA compared to RaTG13 and only PRR to pangolin. Only serine (Ser) in pangolin and both threonine (Thr) and serine (Ser) O-linked glycans were seen in RaTG13.

**Conclusion** Though, pangolin (*Manis javanica*) and bat (*Rhinolophus affinis*) related CoV proximal to SARS-CoV; detailed study needed to confirm.

## Background:

An unknown etiology of acute respiratory disease in late December 2019 was linked to a Huanan seafood wholesale market in Wuhan, China, where over 100 wet animals were on sale before the outbreak [1–3]. World Health Organization (WHO) initially named this strain, novel coronavirus 2019 (2019-nCoV) on 12 January 2020 and officially named the disease as coronavirus disease 2019 (COVID-19) on 11 February 2020 [4]. COVID-19 symptoms are fever, cough, sore throat, fatigue, malaise, breathlessness and it may progress to pneumonia, acute respiratory distress syndrome (ARDS) and multi-organ dysfunction [5]. Up to 14 May 2020, COVID-19 spread has been the highest in Americas, followed by Europe, Eastern Mediterranean, Western Pacific, South-East Asia, Africa and these figures are being updated daily and are expected to increase further [6]. Consumption of wild animals or direct contact with intermediate host animals was suspected to be an initial mode of transmission. However, source transmission routes of SARS-CoV-2 remain obscure [7, 8]. The mystery of intermediate host finding will provide support to prevent further spread, to develop the targeted vaccine and antiviral drugs. The recent studies documents *Rhinolophus affinis*, bat-CoV (RaTG13) and *Manis javanica*, Pangolin-CoV were proximal to SARS-CoV-2 [9–11]. Here, we aimed to update the origin of SARS-CoV-2 by systematic phylogenetic classification and spike glycoprotein (S protein) amino acid sequences.

## Methods:

### Sequences used in the study:

Full-length protein sequences of S protein were downloaded from the NCBI GenBank Database, including eleven wild type SARS-CoV-2 (accession number: QIQ22760, QHS34546, QHR84449, QHZ00379, QIC53204, QHO62877, QHD43416.1, QIA98554, QIK50417, QIK50438 and QIQ08790), SARS-CoV (accession number: AAP13441.1), MERS-CoV (accession number: AFS88936.1), bat-SARS-like-CoV (SL-CoV) (accession number: AVP78042 and AVP7831), bat (*Rhinolophus affinis*) CoV-RaTG13 (accession number: QHR63300.2), Malayan Pangolin (*Manis javanica*) derived CoV (accession number: QIA48641, QIA48614, QIQ54048, QIA48632 and QIA48623.1) and representative viruses of the *Coronavirinae* subfamily (alpha-CoV, beta-CoV, gamma-CoV and Delta-CoV) [12].

### Phylogenetic analysis and Protein sequences alignment:

For phylogenetic analysis, the full length S protein sequences of 11 countries SARS-CoV-2 were compared with SARS-CoV, MERS-CoV, bat-CoV (RaTG13), Pangolin-CoV, bat-SL-CoV and previously published representative viruses of the *Coronavirinae* subfamily sequences by BLAST-EXPLORER program that uses the neighbor-joining method with 1000 bootstrap replicates [13]. The resulting dendrograms were used to verify previously proposed genera assignments and identify areas for clarification. Alignment of RBD and O-linked glycan residues sequences between SARS-CoV-2 strains, RaTG13, Pangolin-CoV, bat-SL-CoV and SARS-CoV, were analyzed by MEGA-10 [14].

## Results:

Efforts to identify the reservoir of human CoV led to the discovery of diverse CoV, which are genetically close related. For the first time, we have constructed an “S” protein sequence-based phylogenetic tree with all the known *Coronavirinae* subfamily viruses for the betterment of understanding of current SARS-CoV-2 clustering and classified them into genera alpha, beta, gamma and delta CoV. To cross-check the proximal to SARS-CoV-2; we had chosen wild type human CoV spike protein sequence to compare with all species of CoV along with recently documented closest CoV (RaTG13 and Pangolin-CoV) [Figure 1]. The protein sequences were nearly identical across the S protein of eleven isolates, with sequence identity above 99.70%, indicative of a very recent emergence into the human population and justification here why we selected those eleven isolates than mutated and variant strains being updated globally. The phylogenetic analysis result showed that eleven SARS-CoV-2 isolates were closely clustered to inner joint neighbor RaTG13 (97.41%), Pangolins carried CoV (92.22% identity) and bat-SL-CoV (80.36% identity). All these together form a new clade 2 in lineage B of beta CoV and 2003 emerged SARS-CoV (Urbani) forms clade 1.

The CoV S protein is an envelope glycoprotein that plays the most important role in viral attachment, fusion, and entry into host cells, and serves as a major target for the development of neutralizing

antibodies, inhibitors of viral entry, and vaccines. SARS-CoV-2 S protein (1273aa) contains two functional domains, such as receptor binding domain (RBD) (223aa sequence 319–541 aa) and glycoprotein domain (609aa sequence 662–1270 aa) [15]. In this study, SARS-CoV-2, RaTG13, Pangolin-CoV and Bat-SL-CoV sequence were analyzed for protein function from the Conserved Domain Database in NCBI [16] (**Supplement-1**). First isolate of SARS-CoV-2 (QHD43416), RaTG13 and 5 Pangolin CoV were grouped under CD21480 protein family [17], whereas Bat-SL-CoV (234aa; 326–560) and SARS-CoV grouped in c109656 and pfam09408, respectively [18]. Interestingly, the SARS-CoV-2 RBD sequence from 319 to 541 possesses 100% identity in all 11 isolates except one mutation in Indian strain at 408 residue arginine (R) is replaced by isoleucine (I). 90.13% (201/223) identical amino acid sequences were seen between SARS-CoV-2 and RaTG13, while 86.10% (192/223), 66.83% (149/223), 73.09% (163/223) seen in SARS-CoV-2 V s pangolin-CoV, SARS-like-CoV, SARS-CoV, respectively. This investigation showed that RaTG13 and pangolin derived CoV closely related to SARS-CoV-2 than SL-CoV and SARS-CoV (**Supplement-2**).

Study demonstrates that two residue changes between civet and human ACE2, T31K and E38D, present major species barriers between the two species for SARS-CoV infections and we notices E38D in pangolin not in bat [19] [**Figure-2**]. Six RBD amino acids (L455, F486, Q493, S494, N501 and Y505) are critical for binding to ACE2 receptors and for determining the host range of SARS-CoV-2 [11, 19, 20]. Five and four of these six residues differ between SARS-CoV-2 V s RaTG13 and Pangolin-CoV, respectively. Whereas a pangolin-CoV (project ID: PRJNA573298) had 100% (6/6) residues similar to SARS-CoV-2, suggesting there are differences among pangolin-CoV and 50% (3/6) similarity were seen between RaTG13 and Pangolin-CoV [**Figure-3A**]. Another notable feature of SARS-CoV-2 is a PRRA insertion compared to RaTG13 and only PRR to pangolin-CoV in a polybasic cleavage site. This polybasic cleavage site in S2 glycoprotein (681to 684) has a role in determining viral infectivity and host range. Besides, O-linked glycans to S673, T678 and S686, which side the cleavage site of SARS-CoV-2 and compared to RaTG13 and Pangolin-CoV. Only serine (Ser) in pangolin and both threonine (Thr) and serine (Ser) O-linked glycans were seen in RaTG13. Between the O-linked glycans, pangolin had varied aa sequences compared to RaTG13 [**Figure-3B**].

## Discussion:

CoV are enveloped have a non-segmented, positive-sense RNA genome ranging from 26 to 32-kilo bases in length [21] and divided into four genera, including Alpha/Beta/Delta/Gamma. Evolutionary analyses have shown that bats, civet, camel, murine, canine, bovine, equine and rodents are the gene sources of most alpha-CoV and beta-CoV, while avian species, whale and porcine are the gene sources of most delta-CoV and gamma-CoV [22, 23]. Prior to December 2019, 6 CoV were described to be pathogenic to humans [24]. In this study, for the first time we have constructed the phylogenetic tree with all the species of the CoV and current outbreak of SARS-CoV-2, seventh human CoV infection belong to beta-CoV [lineage B].

The earliest genomic characterization of SARS-CoV-2 strains in Wuhan had 88–89% nucleotide identity with bat-SL-CoV (bat-SL-CoVZC45 and bat-SL-CoVzxc21), 79–89% nucleotide identity with human SARS-CoV and more distant from MERS-CoV (50%) [1, 23, 25, 26]. Although SARS-CoV-2 epidemic was linked to

Wuhan seafood market, Huang C, et al. reported a total of 41 patients, and 14 cases are not related to the seafood market and no trace of bats has been found, so exact place of origin need to be studied in detail [27]. Subsequently, Zhou P. et al. from Wuhan institute of virology (Zheng Li Shi lab) showed that SARS-CoV-2 was highly similar throughout the genome to RaTG13 with an overall genome sequence identity of 96.2% and 93.1% nucleotide identity to S protein. Also, the author did not mention when it has been sequenced and RNA dependent RNA polymerase (RdRp) data not shown to compare SARS-CoV-2 [9]. RaTG13 was isolated from the bat (*Rhinolophus affinis*) on 24 July 2013 by Zheng Li Shi group and the reason unclear why they did not submit the sequence before instead on 27 Jan 2020, although it is proximal to bat-SL-CoV (accession number: AVP78042.1, AVP78031.1 and ACU31051.1) (**Supplement-3**). SARS-CoV (Rs806/2006) (accession number: ACU31051.1) already has proven for Intraspecies diversity and its implications for the origin of SARS coronaviruses in humans [28]. Hence, the detailed investigation needed for RaTG13 isolate and origin. Scientists report genetic sequences of viruses isolated from pangolins are 99% similar to that of the COVID-19 strains [7, 8, 10, 29]. Lam TT, et al. identified two sub-lineages of SARS-CoV-2-related CoV in Malayan pangolin, one that exhibits strong similarity to SARS-CoV-2 in the RBD [30]. Zhang C, et al., assembled a draft genome of the SARS-CoV-2 using the metagenomic samples from the lung of *Manis javanica*, showing an overall coverage of 73% of COVID-19 strains with 91% sequence identity [31]. However, Li X, et al. concluded that the human SARS-CoV-2 virus, did not come directly from pangolins based on a unique peptide (PRRA) insertion seen in the human SARS-CoV-2 virus and not in pangolins carried CoV [32]. Also, a study demonstrated SARSCoV-2 is not a purposefully manipulated virus, based on high-affinity binding to human ACE2, polybasic cleavage site and the three adjacent predicted O-linked glycans are unique to SARS-CoV-2 and were not previously seen in lineage B beta-CoV [11]. Hence, we compared RaTG13 and pangolin-CoV with SARS-CoV-2 for an update and betterment of understanding.

RBD of S protein in SARS-CoV-2 binds strongly to human, pangolin and bat angiotensin-converting enzyme 2 (ACE2) receptors [19, 20, 33]. Studies have confirmed that S protein in the SARS-CoV-2 uses the ACE2, found in the lower respiratory tract of humans [1, 9], and other certain species (pangolin, civet, swine, cow, buffalo, goat, cat, sheep and pigeon) as cellular entry receptor [34, 35]. Liu Z, et al. indicated that, other than pangolins and snakes, turtles may act as the potential intermediate hosts transmitting SARS-CoV-2 to humans based on the key amino acid interaction between RBD and ACE2 [36]. Choudhury A, et al. showed SARS-CoV-2 is close to bat-CoV, strongly binds with ACE2 receptor protein from both human and bat origin and TLR4 is most likely to be involved in recognizing molecular patterns from SARS-CoV-2 to induce inflammatory responses [37]. A study data support the natural origin of SARS-CoV-2, likely derived from bats, possibly transferred to pangolins, before spreading to man and it not artificial CoV, including the chimeric SL-SHC014-MA15 [38]. The study proposes a unique cleavage motif promoting SARS-CoV-2 infection in humans may be under strong selective pressure, given that replication in permissive Vero-E6 cells leads to the loss of this adaptive function [39]. Overall, we demonstrate the key residues of RBD (455, 486, 493, 494, 501 and 505) and polybasic cleavage sites varies significantly; need to be studied in detail for a better understanding of cross-species transmission. PubMed search

result showed only three bats (*Rhinolophus affinis*) and five pangolin CoV sequences were available and more CoV isolation needs to verify the origin of RaTG13.

## **Conclusion:**

Although RaTG13 and Pangolin derived CoV are very proximal to SARS-CoV-2, the key receptor binding and O-linked glycan residues vary significantly, except a Malayan pangolin (PRJNA573298) isolate has 100% identity. The Polybasic cleavage site (PRRA insertion) was absent in RaTG13 and pangolin (PRJNA573298), whereas it is only PRR in other pangolin isolates with unique amino acid changes within. Thus, animal study, isolation of CoV from pangolin (*Manis javanica*) and bat (*Rhinolophus affinis*) is necessary to help in understanding of SARS-CoV-2 origin and intermediate transmission.

## **List Of Abbreviations:**

SARS-CoV-2 - Acute respiratory syndrome coronavirus-2

ACE2 - Angiotensin-converting enzyme 2

RBD - Receptor-binding domain

WHO - World Health Organization

2019-nCoV - Novel coronavirus 2019

COVID-19- coronavirus disease 2019

ARDS - Acute respiratory distress syndrome

## **Declarations**

## **Funding information:**

The author received no specific grant from any funding agency.

## **Conflicts of interest:**

The author declares that there are no conflicts of interest.

## **Ethical approval and Consent for publication:**

NA

## Availability of data and material:

Yes

## Code availability:

N/A

## Consent for publication:

Yes.

## Author's contributions:

Jeevan Malaiyan, contributed to the conceptualization, study design, critical review of the content and approved the final version of the manuscript.

Suresh Arumugam, contributed to study design, data analysis and approved the final version of the manuscript.

Kamalraj Mohan, contributed to data analysis and approved the final version of the manuscript.

Gokul Gomathi Radhakrishnan, contributed to data analysis and approved the final version of the manuscript.

## Acknowledgments:

Authors thank the global doctors and scientists who identified the SARS-CoV-2, RaTG13, Pangolin-CoV and related gene sequences.

## References

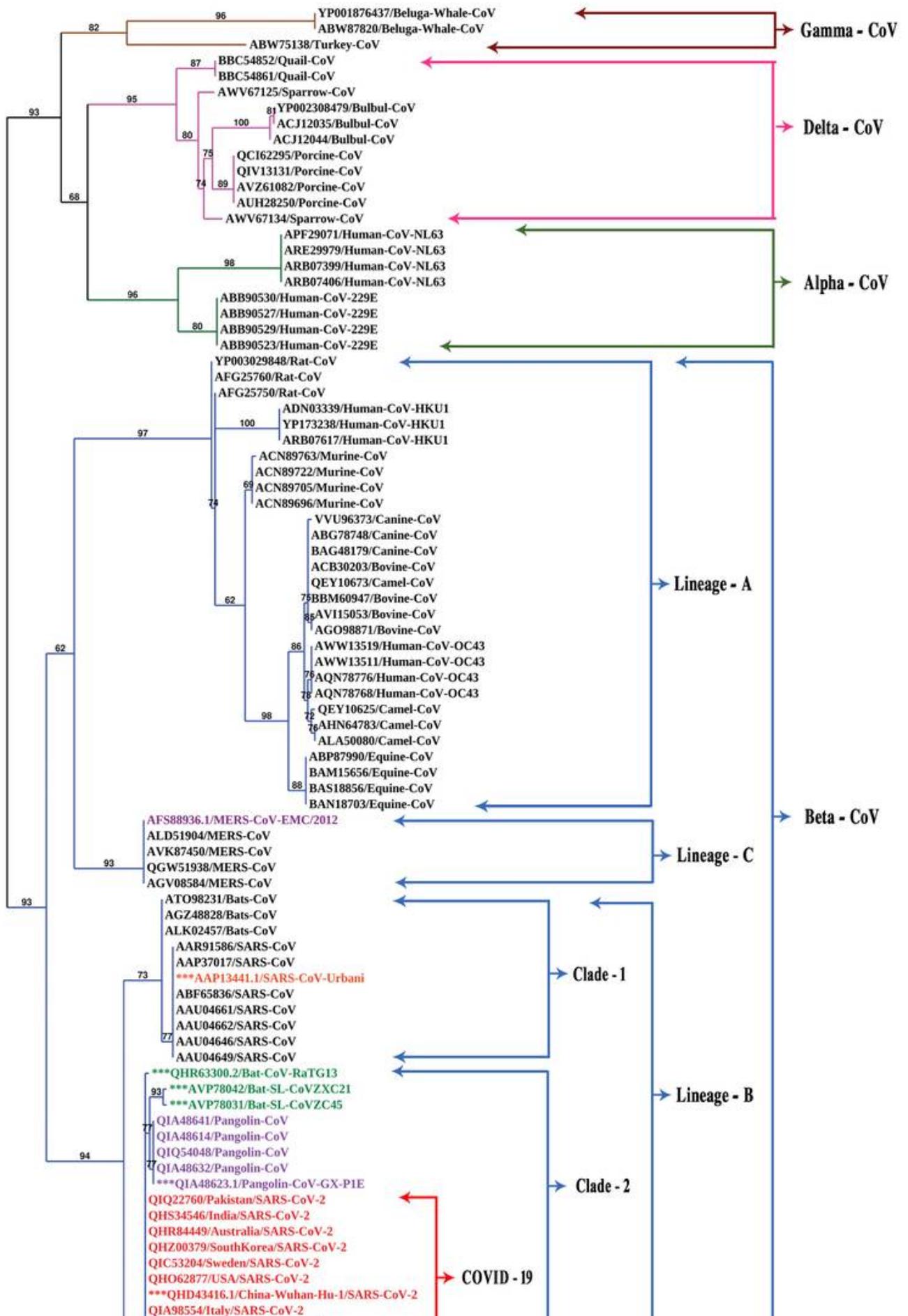
1. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W et. al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395(10224):565-574.
2. Jung SM, Kinoshita R, Thompson RN, Linton NM, Yang Y, Akhmetzhanov AR et al. Epidemiological Identification of A Novel Pathogen in Real Time: Analysis of the Atypical Pneumonia Outbreak in Wuhan, China, 2019-2020. *Journal of Clinical Medicine*. 2020;9(3):637.
3. Zhang YZ, Holmes EC. A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell*. 2020; <https://doi:10.1016/j.cell.2020.03.035>

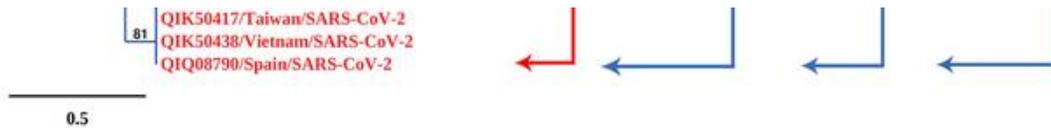
4. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Mil Med Res.* 2020;7(1):11.
5. Singhal T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J Pediatr*2020;87:281-286.
6. World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int>. Accessed 14 May 2020.
7. Yang Y, Peng F, Wang R, Guan K, Jiang T, Xu G et al. The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. *Journal of autoimmunity.* 2020;109:102434.
8. Liu P, Chen W, Chen JP. Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (*Manis javanica*). *Viruses.* 2019;11(11):1-15.
9. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature.* 2020;579(7798):270-3.
10. Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr Biol.* 2020;30(7):1346-1351.
11. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med.* 2020;26(4):450-452.
12. Shafique L, Ihsan A, Liu Q. Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. *Pathogens.* 2020;9(3):240.
13. Dereeper A, Audic S, Claverie JM, Blanc G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC evolutionary biology.* 2010;10(1):8.
14. Kumar S, Stecher G, Li M, Knyaz, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 2018;35:1547-1549.
15. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 2017;45(D1):D200-D203.
16. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. "CDD/SPARCLE: the conserved domain database in 2020.", *Nucleic Acids Res.* 2020;48(D1):D265-D268.
17. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020;581(7807):215-220.
18. Prabakaran P, Gan J, Feng Y, Zhu Z, Choudhry V, Xiao X et al. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J Biol Chem.* 2006;281:15829-15836.
19. Li F. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral research.* 2013; 100(1): 246-54.
20. Jia HP, Look DC, Shi L, Hickey M, Pewe L, Netland J, et al. ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia.

- Journal of virology. 2005;79(23):14614-21.
21. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol* 2016;24:490-502.
  22. Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology*. 2018; 23(2):130-7.
  23. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727–33.
  24. Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol*. 2005;79(2):884–895.
  25. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269.
  26. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect*. 2020;9(1):221-236.
  27. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*. 2020 Feb 15;395(10223):497-506.
  28. Yuan J, Hon CC, Li Y, Wang D, Xu G, Zhang H, Zhou P, Poon LL, Lam TT, Leung FC, Shi Z. Intraspecies diversity of SARS-like coronaviruses in *Rhinolophus sinicus* and its implications for the origin of SARS coronaviruses in humans. *Journal of general virology*. 2010 Apr 1;91(4):1058-62.
  29. Cyranoski D. Mystery deepens over animal source of coronavirus. *Nature*. 2020;579(7797):18-19.
  30. Lam TT, Shum MH, Zhu HC, Tong YG, Ni XB, Liao YS, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*. 2020 Mar 26. doi: 10.1038/s41586-020-2169-0
  31. Zhang C, Zheng W, Huang X, Bell EW, Zhou X, Zhang Y. Protein Structure and Sequence Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique Similarity between Its Spike Protein Insertions and HIV-1. *J Proteome Res*. 2020; 19(4):1351-1360.
  32. Li X, Zai J, Zhao Q, Nie Q, Li Y, Foley BT, Chaillon A. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV- Journal of medical virology. 2020; [https://doi: 10.1002/jmv.25731](https://doi.org/10.1002/jmv.25731)
  33. Tai W, He L, Zhang X, Jing Pu, Voronin D, Jiang S, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol* (2020). <https://doi.org/10.1038/s41423-020-0400-4>
  34. Qiu Y, Zhao YB, Wang Q, Li JY, Zhou ZJ, Liao CH et al. Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2. *Microbes Infect*. 2020; S1286-4579(20)30049-6. doi: 10.1016/j.micinf.2020.03.003.

35. Luan J, Lu Y, Jin X, Zhang L. Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochem Biophys Res Commun*. 2020; doi: 10.1016/j.bbrc.2020.03.047
36. Liu Z, Xiao X, Wei X, Li J, Yang J, Tan H, Zhu J, et al. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J Med Virol*. 2020 Feb 26. doi: 10.1002/jmv.25726.
37. Choudhury A, Mukherjee S. In silico studies on the comparative characterization of the interactions of SARS-CoV-2 spike glycoprotein with ACE-2 receptor homologs and human TLRs. *Journal of Medical Virology* 2020; doi: 10.1002/jmv.25987
38. Dallavilla T, Bertelli M, Morresi A, Bushati V, Stuppia L, Beccari T, et al. Bioinformatic analysis indicates that SARS-CoV-2 is unrelated to known artificial coronaviruses. *Eur Rev Med Pharmacol Sci*. 2020;24(8):4558-4564.
39. Lau SY, Wang P, Mok BW, Zhang AJ, Chu H, Lee AC, Deng S, Chen P, Chan KH, Song W, Chen Z. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerging Microbes & Infections*. 2020; doi: 10.1080/22221751.2020.1756700.

## Figures





**Figure 1**

Phylogenetic analysis of S protein of SARS-CoV-2 strains and representative viruses of the Coronavirinae subfamily. Countrywide first reported SARS-CoV-2 isolates were closely clustered to RaTG13 (97.41% identity), Pangolins-CoV (92.22% identity) and bat-SL-CoV (80.36% identity) forms a new clade 2 in lineage B of beta CoV.

**A**

ACE2	31	35	38	41	82	353
Human	K	E	D	Y	M	K
Civet	T	E	E	Y	T	K
Pangolin	K	E	E	Y	N	K
Bat	N	K	D	H	T	K
Mouse	H	E	D	Y	S	H

**B**

RBD	Year	455	486	493	494	501	505
SARS-CoV-2	2019	L	F	Q	S	N	Y
Bat CoV-RaTG13	2013	L	L	Y	R	D	H
Pangolin-CoV	2017	L	L	E	R	T	Y
hSARS-CoV	2003	Y	L	N	D	T	Y
SL-CoV	2015	S	F	S	T	V	Y

**Figure 2**

Receptor and RBD residues that play important roles in SARS-CoV-2: (A) Alignment of ACE2 residues from different animal species that is critical for the host range of SARS-CoV and SARS-CoV-2. The GenBank accession numbers are AY623811 (human ACE2), AY881174 (civet ACE2), XP\_017505752.1 (Pangolin), EF408740 (mouse ACE2), and GQ999937 (bat ACE2). (B) Alignment of RBD residues from different SARS-CoV-2. Figure adapted from [19, 20].

