Received: 14 May 2020 | Accepted: 4 July 2020

DOI: 10.1002/jmv.26261

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

An update on the origin of SARS-CoV-2: Despite closest identity, bat (RaTG13) and pangolin derived coronaviruses varied in the critical binding site and O-linked glycan residues

Jeevan Malaiyan[1] | Suresh Arumugam[2] | Kamalraj Mohan[1] | Gokul Gomathi Radhakrishnan[3]

[1]Division of Virology, Department of Microbiology, Sri Muthukumaran Medical College Hospital and Research Institute, Affiliated to The Tamil Nadu Dr. M.G.R. Medical University, Chikkarayapuram, Chennai, India

[2]Central Research Facility, Meenakshi Medical College Hospital and Research Institute, Meenakshi Academy of Higher Education and Research, Kancheepuram, Tamilnadu, India

[3]Department of General Medicine, Sri Muthukumaran Medical College Hospital and Research Institute, Affiliated to The Tamil Nadu Dr. M.G.R. Medical University, Chikkarayapuram, Chennai, India

Correspondence
Jeevan Malaiyan, Division of Virology, Department of Microbiology, Sri Muthukumaran Medical College Hospital and Research Institute, Affiliated to The Tamil Nadu Dr. M.G.R. Medical University, Chikkarayapuram 600 069, Chennai, India.
Email: jeevan1209@gmail.com

Abstract

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 23 June 2020 with 8 993 659 active cases, 469 587 deaths across 216 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. 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) and O-linked glycan were compared. 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 PRRA 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, suggesting that a detailed study needed in pangolin (*Manis javanica*) and bat (*Rhinolophus affinis*) related CoV.

KEYWORDS

COVID-19, intermediate host, pangolins, RaTg13, SARS-CoV-2

1 | INTRODUCTION

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 23 June 2020, COVID-19 spread has been the highest in America, followed by Europe, Eastern Mediterranean, South-East Asia, Africa, and Western Pacific 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 severe acute respiratory

syndrome coronavirus-2 (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.

2 | MATERIALS AND METHODS

2.1 | 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, QIK 50438, 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 (α-CoV, β-CoV, γ-CoV, and Δ-CoV).[12]

2.2 | 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]

3 | 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 α, β, γ, and Δ 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 11 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 β 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-2S 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] (Figure S1). First isolate of SARS-CoV-2 (QHD43416), RaTG13, and five pangolin CoV were grouped under the 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 vs 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 (Figure S2).

ACE2 receptor and RBD residues play important roles in SARS-CoV-2 entry. 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] Five and four of these six residues differ between SARS-CoV- 2 Vs 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 2A]. 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. The functional consequence of the polybasic cleavage site in SARS-CoV-2 is unknown, and it will be important to determine its impact on transmissibility and pathogenesis in animal models. 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 2B) and the structural difference of ACE2 contact residues and O-linked glycan among SARS-CoV are shown in Table 1.

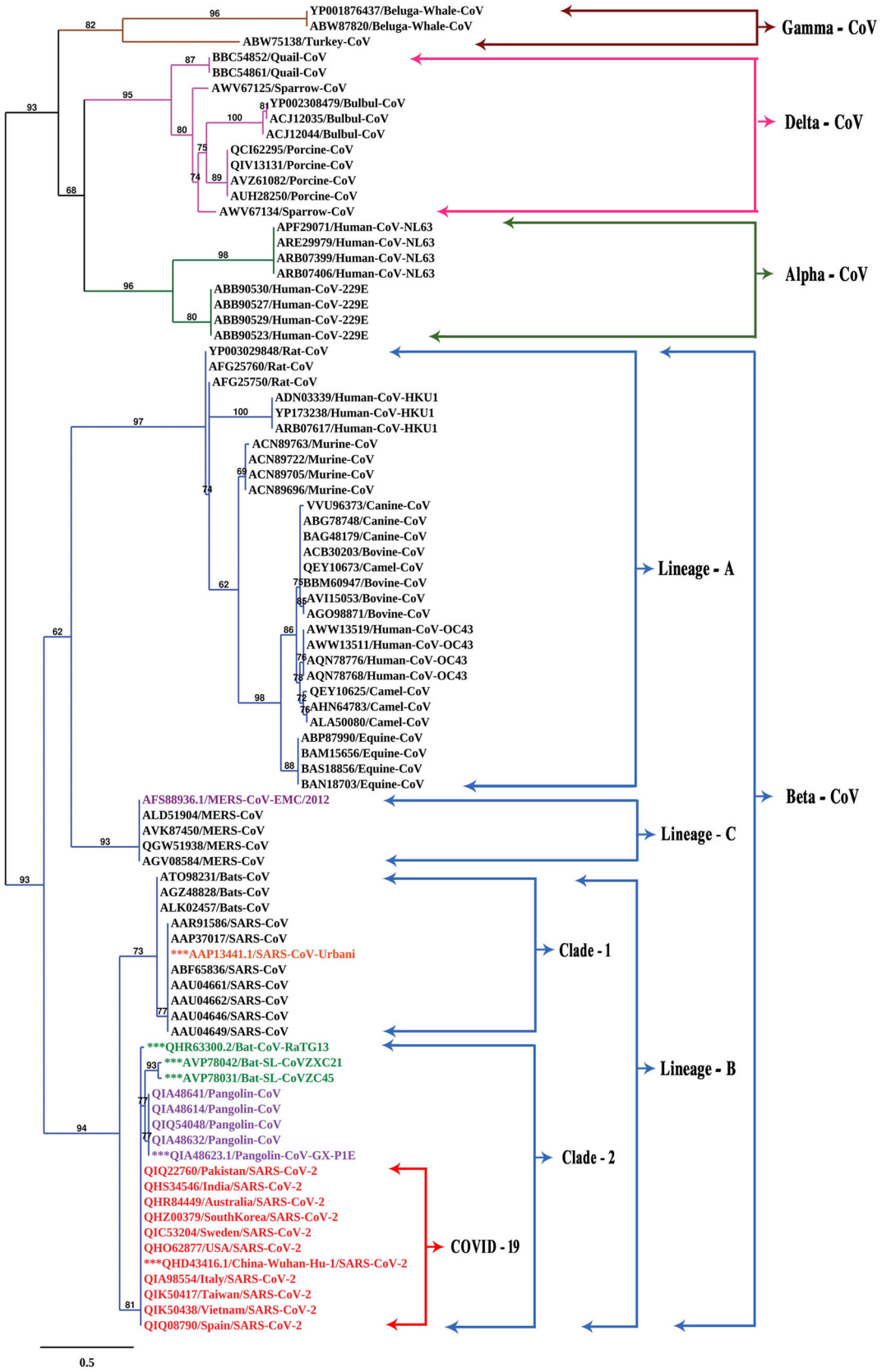


FIGURE 1 Phylogenetic analysis of S protein of SARS-CoV-2 strains and representative viruses of the *Coronavirinae* subfamily. Countrywide first reported SARS-CoV-2 isolates were closely clustered to RaTG13 (97.41% identity), pangolins-CoV (92.22% identity), and bat-SL-CoV (80.36% identity) forms a new clade 2 in lineage B of β CoV. SARS-CoV-2, severe acute respiratory syndrome coronavirus-2

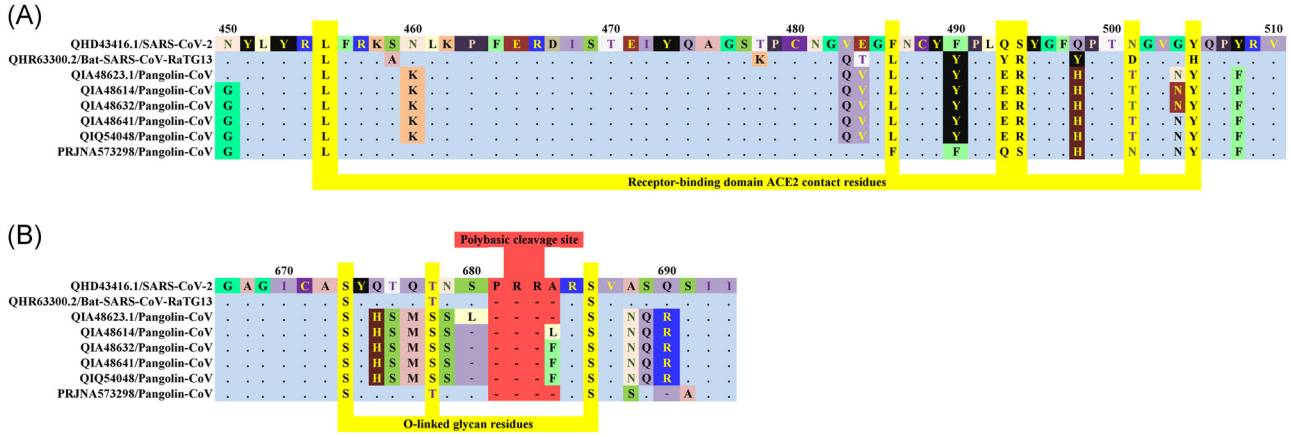


FIGURE 2 Features of the spike protein in human SARS-CoV-2, RaTG13 and pangolin-CoV. A, Mutations in contact residues of the SARS-CoV-2 S protein. Key residues in the spike protein that make contact to the ACE2 receptor are marked with yellow in both SARS-CoV-2 and related viruses. B, Acquisition of polybasic cleavage site (Red) and O-linked glycans (yellow). SARS-CoV-2, severe acute respiratory syndrome coronavirus-2

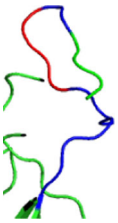
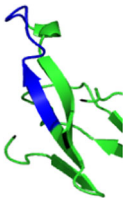
4 | DISCUSSION

CoV are enveloped have a nonsegmented, positive-sense RNA genome ranging from 26 to 32-kilo bases in length[20] and divided into four genera, including α/β/Δ/γ. Evolutionary analyses have shown that bats, civet, camel, murine, canine, bovine, equine, and rodents are the gene sources of most α-CoV and β-CoV, whileavian species, whale and porcine are the gene sources of most Δ-CoV and γ-CoV.[21,22] Before December 2019, 6 CoV were described to be pathogenic to humans.[23] 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 β-CoV (lineage B).

The earliest genomic characterization of SARS-CoV-2 strains in Wuhan had 88% to 89% nucleotide identity with bat-SL-CoV (bat-SL-CoVZC45 and bat-SL-CoVzxc21), 79% to 89% nucleotide identity with human SARS-CoV and more distant from MERS-CoV (50%).[1,22,24,25] Although the SARS-CoV-2 epidemic was linked to the Wuhan seafood market, Huang et al[26] 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.[26] Subsequently, Zhou et al[9] 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 January 2020, although it is proximal to bat-SL-CoV (accession number: AVP78042.1, AVP78031.1, and ACU31051.1) (Figure S3). 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.[27] 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.[7,8,10,28] Lam 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.[29] Zhang 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.[30] However, Li et al[31] 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 β-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.[19,33,34] 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.[35,36] Liu et al[37] 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. Choudhury et al[38] 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. Li et al[39] demonstrated that SARS-CoV-2's entire RBM was introduced through recombination with CoV from pangolins, possibly a critical step in the evolution of SARS-CoV-2's ability to infect humans. 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.[40] The study proposes a unique cleavage motif promoting SARS-CoV-2 infection in humans may be under strong selective pressure, given that replication in

TABLE 1 3D structural difference found in receptor-binding domain ACE2 contact residues and O-linked glycan residues among SARS-CoV

Strain name	Receptor-binding domain ACE2 contact residues	O-linked glycan residues
QHD43416.1/SARS-CoV-2		 [a]
QHR63300.2/RaTG13		
QIA48623.1/pangolin-CoV		
AVP78031.1/bat-SARS-CoV		
AAP13441.1/SARS-CoV-Urbani		

Abbreviations: ACE-2, angiotensin-converting enzyme 2; SARS-CoV-2, severe acute respiratory syndrome coronavirus-2
[a] Red color loop showed the PRRA site in O linked glycan residues.

permissive Vero-E6 cells leads to the loss of this adaptive function.[41] 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 results showed only three bat (*Rhinolophus affinis)* and five pangolin CoV sequences were available and more CoV isolation need to verify the origin of RaTG13.

5 | CONCLUSION

Although RaTG13 and pangolin-derived CoV is 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 the understanding of SARS-CoV-2 origin and intermediate transmission.

ACKNOWLEDGMENTS

The authors thank the global doctors and scientists who identified the SARS-CoV-2, RaTG13, pangolin-CoV, and related gene sequences.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

JM contributed to the conceptualization, study design, critical review of the content and approved the final version of the manuscript. SA contributed to study design, data analysis, and approved the final version of the manuscript. KM contributed to data analysis and approved the final version of the manuscript. GGR contributed to data analysis and approved the final version of the manuscript.

ORCID

Jeevan Malaiyan http://orcid.org/0000-0001-6466-308X
Suresh Arumugam http://orcid.org/0000-0001-6247-1156

REFERENCES

1. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395(10224):565-574. https://doi.org/10.1016/S0140-6736(20)30251-8
2. Jung SM, Kinoshita R, Thompson RN, et al. Epidemiological identification of a novel pathogen in real time: analysis of the atypical pneumonia outbreak in Wuhan, China, 2019-2020. *J Clin Med*. 2020;9(3):637. https://doi.org/10.3390/jcm9030637
3. Zhang YZ, Holmes EC. A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell*. 2020;181:223-227. https://doi.org/10.1016/j.cell.2020.03.035
4. Guo YR, Cao QD, Hong ZS, 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. https://doi.org/10.1186/s40779-020-00240-0
5. Singhal T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J Pediatr*. 2020;87:281-286.
6. WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int. Accessed on 23 May 2020.
7. Yang Y, Peng F, Wang R, et al. The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. *J Autoimmun*. 2020;109:102434. https://doi.org/10.1016/j.jaut.2020.102434
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. https://doi.org/10.3390/v11110979
9. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273. https://doi.org/10.1038/s41586-020-2012-7
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. https://doi.org/10.1016/j.cub.2020.03.022
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. https://doi.org/10.1038/s41591-020-0820-9
12. Rehman S, Shafique L, Ihsan A, Liu Q. Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. *Pathogens*. 2020;9(3):240. https://doi.org/10.3390/pathogens9030240
13. Dereeper A, Audic S, Claverie JM, Blanc G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol*. 2010;10(1):8. https://doi.org/10.1186/1471-2148-10-8
14. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547-1549. https://doi.org/10.1093/molbev/msy096
15. Marchler-Bauer A, Bo Y, Han L, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. 2017;45(D1):D200-D203. https://doi.org/10.1093/nar/gkw1129
16. Lu S, Wang J, Chitsaz F, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 2020;48(D1):D265-D268. https://doi.org/10.1093/nar/gkz991
17. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;581:215-220. https://doi.org/10.1038/s41586-020-2180-5
18. Prabakaran P, Gan J, Feng Y, et al. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J Biol Chem*. 2006;281:15829-15836. https://doi.org/10.1074/jbc.M600697200
19. Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature*. 2020;581(7807):221-224. https://doi.org/10.1038/s41586-020-2179-y
20. Su S, Wong G, Shi W, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol*. 2016;24:490-502. https://doi.org/10.1016/j.tim.2016.03.003
21. Yin Y, Wunderink RGMERS. SARS and other coronaviruses as causes of pneumonia. *Respirology*. 2018;23(2):130-137. https://doi.org/10.1111/resp.13196
22. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727-733. https://doi.org/10.1056/NEJMoa2001017
23. Woo PC, Lau SK, Chu CM, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol*. 2005;79(2):884-895. https://doi.org/10.1128/JVI.79.2.884-895.2005
24. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. https://doi.org/10.1038/s41586-020-2008-3
25. Chan JFW, Kok KH, Zhu Z, 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. https://doi.org/10.1080/22221751.2020.1719902
26. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*. 2020;395(10223):497-506. https://doi.org/10.1016/S0140-6736(20)30183-5
27. Yuan J, Hon CC, 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 Gen Virol*. 2010;91(4):1058-1062. https://doi.org/10.1099/vir.0.016378-0
28. Cyranoski D. Mystery deepens over animal source of coronavirus. *Nature*. 2020;579(7797):18-19. https://doi.org/10.1038/d41586-020-00548-w
29. Lam TT, Jia N, Zhang YW, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*. 2020;583:282–285. https://doi.org/10.1038/s41586-020-2169-0
30. 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. https://doi.org/10.1021/acs.jproteome.0c00129
31. Li X, Zai J, Zhao Q, et al. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. *J Med Virol*. 2020;92:602-611. https://doi.org/10.1002/jmv.25731
32. Liu P, Jiang JZ, Wan XF, et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLOS Pathog*. 2020;16(5):e1008421. https://doi.org/10.1371/journal.ppat.1008421
33. Tai W, He L, Zhang X, 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;17(6):613–620. https://doi.org/10.1038/s41423-020-0400-4

34. Zhai X, Sun J, Yan Z, et al. Comparison of SARS-CoV-2 spike protein binding to ACE2 receptors from human, pets, farm animals, and putative intermediate hosts. *J Virol*. 2020. https://doi.org/10.1128/JVI.00831-20
35. Qiu Y, Zhao YB, Wang Q, et al. Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2. *Microbes Infect*. 2020;22(4-5):221–225. https://doi.org/10.1016/j.micinf.2020.03.003
36. 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;526:165-169. https://doi.org/10.1016/j.bbrc.2020.03.047
37. Liu Z, Xiao X, Wei X, 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;92:595-601. https://doi.org/10.1002/jmv.25726
38. 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. *J Med Virol*. 2020: 1–9. https://doi.org/10.1002/jmv.25987
39. Li X, Giorgi EE, Marichannegowda MH, et al. Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv*. 2020;6(27):eabb9153. https://doi.org/10.1126/sciadv.abb9153
40. Dallavilla T, Bertelli M, Morresi A, 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. https://doi.org/10.26355/eurrev_202004_21041
41. Lau SY, Wang P, Mok BW, et al. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerging Microbes & Infections*. 2020;9:837-842. https://doi.org/10.1080/22221751.2020.1756700

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Malaiyan J, Arumugam S, Mohan K, Gomathi Radhakrishnan G. An update on the origin of SARS-CoV-2: Despite closest identity, bat (RaTG13) and pangolin derived coronaviruses varied in the critical binding site and O-linked glycan residues. *J Med Virol*. 2021;93: 499–505. https://doi.org/10.1002/jmv.26261