Molecular evolution and phylogenetic analysis of SARS-CoV-2 and hosts ACE2 protein suggest Malayan pangolin as intermediary host

Luciano Rodrigo Lopes [1] · Giancarlo de Mattos Cardillo [2] · Paulo Bandiera Paiva [1]

Abstract

An emergence of a novel coronavirus, causative agent of COVID19, named as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), occurred due to cross-species transmission. Coronaviruses are a large family of viruses able to infect a great number of hosts. Entrance of SARS-CoV-2 depends on the surface (S) protein interaction with host ACE2 protein and cleavage by TMPRSS2. ACE2 could be a species-specific barrier that interferes with bat-to-human coronavirus cross-species transmission. Molecular analysis supported bats as natural hosts for SARS-CoV and involved them in MERS-CoV origin. The genomic similarity between bat RaTG13 CoV strain and SARS-CoV-2 implicates bats in the origin of the new outbreak. Additionally, there is a hypothesis for the zoonotic transmission based on contact with Malayan pangolins by humans in Huanan seafood market in Wuhan, China. To investigate bats and pangolin as hosts in SARS-CoV-2 cross-species transmission, we perform an evolutionary analysis combining viral and host phylogenies and divergence of ACE2 and TMPRSS2 amino acid sequences between CoV hosts. Phylogeny showed SARS-like-CoV-2 strains that infected pangolin and bats are close to SARS-CoV-2. In contrast to TMPRSS2, pangolin ACE2 amino acid sequence has low evolutionary divergence compared with humans and is more divergent from bats. Comparing SARS-CoV with SARS-CoV-2 origins, pangolin has yet lower ACE2 evolutionary divergence with humans than civet—the main intermediary host of SARS-CoV. Thus, pangolin has become an opportune host to intermediates bat-to-human SARS-CoV-2 jump and entry.

Keywords Coronavirus · SARS-CoV-2 · ACE2 · Host · Bat · SARS-CoV

Global attention has been given to the emergence of a new coronavirus pandemic. In December 2019, a serious pneumonia outbreak caused by a novel coronavirus started in China [1]. The given name to the disease associated with the coronavirus was COVID19, while novel CoV was named as

✉ Luciano Rodrigo Lopes
  luciano.lopes@unifesp.br

[1] Bioinformatics and Biomedical Data Science Division, Health Informatics Department, Universidade Federal de São Paulo–UNIFESP, São Paulo, SP, Brazil

[2] Department of Neurology and Neurosurgery, Universidade Federal de São Paulo–UNIFESP, São Paulo, SP, Brazil

severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). COVID19 is milder but a highly transmissible infectious disease compared with SARS (severe acute respiratory syndrome) and MERS (Middle East respiratory syndrome) outbreaks, according to morbidity and mortality rates [2, 3]. Coronaviruses are a large family of viruses able to infect a great number of hosts [4]. Cross-species transmission of zoonotic coronaviruses (CoVs) can result in disease outbreaks [5]. Molecular analysis supported bats as natural hosts for SARS-CoV, but palm civets (*Paguma larvata*) had a critical role in the transmission to humans [6, 7]. Camels were identified as the natural host for MERS-CoV, despite implication of bats in the origin of this virus [8]. Bats are also implicated in SARS-CoV-2 origin. A very similar SARS-CoV-2 strain (RaTG13 CoV) was detected in *Rhinolophus affinis* bat with 96% genome similarity compared with SARS-CoV-2 genome sequence. Considering that bats were in hibernation when the outbreak occurred, the virus is more likely to have been transmitted via other species [9]. The hypothesis for the zoonotic

transmission route was constructed based on contact with Malayan pangolins (*Manis javanica*) by visitors of Huanan seafood market in Wuhan, China [10]. The linkage with Huanan market was done due to few cases of pneumonia in the city of Wuhan associated with the SARS-CoV-2 [11]. Today, the involvement of the Huanan market with COVID19 origin seems controversial [12, 13]. However, molecular epidemiologic studies also confirmed animal handlers as the earliest infected with SARS-CoV in 2002–2003 outbreaks [7]. The presence of bats and bat products in food and traditional medicine markets occurs due persistent viral infection without clinical symptoms [14]. Contact between bats and wild animals may occur because bats live among a large number of different animal species that is also commercialized in animal markets, giving opportunity to coronavirus transmission [7]. Furthermore, other wild animals can be infected by coronavirus after eating partially digested fruits by bats that were spitted out and fall to the ground with residual bat saliva or part of insects discarded by infected bats [15]. These routes allow viral infection by a diverse set of animals, such as palm civets and Malayan pangolins involved in SARS-CoV and SARS-CoV-2 origins, respectively [7, 10, 15]. In humans, SARS-CoV and SARS-CoV-2 are rapidly spread by respiratory droplets, airborne routes, or direct contact [16]. While CoV related with cold is limited to upper respiratory tract infection, SARS-CoV invades pneumocytes in the lower respiratory tract (lungs) [17, 18]. The first viral interaction with host cells occurs when receptor-binding domains (RBDs) from viral spike or surface (S) protein attach ACE2 (angiotensin-converting enzyme 2) host receptor [6]. Expression of ACE2 appears to be essential for SARS-CoV/ SARS-CoV-2 infection in airway epithelia [18, 19]. Subsequently, SARS-CoV/SARS-CoV-2 employ TMPRSS2 (transmembrane protease serine 2) for cell host entry thanks to its cleavage activity [19]. TMPRSS2 is known to cleave and activate the S protein of SARS-CoV [20] and also plays a role in the spread and immunopathology of SARS-CoV and MERS-CoV in the airways [21].

Receptor recognition for SARS-CoV is considered one of the main barriers between animal species and humans [22]. High variation in the SARS-CoV and SARS-CoV-2 S proteins seems to be essential to lead animal-to-human transmission to human-to-human transmission [23]. To investigate bats and pangolin as hosts in SARS-CoV-2 jump to human, we performed evolutionary analysis based on viral and host molecular phylogenies and also evolutionary divergent pairwise analysis. Based on 87 amino acid sequences from CoV S protein, we inferred a maximum likelihood phylogenetic analysis (Fig. 1), using PhyML 3.3v software to phylogenetic tree inference and MEGA X to find the best substitution model (Supplementary Table 1). Rooted by cold-related 229E CoV, the phylogenetic tree formed three distinct clades. Clade 1 includes two main clusters: clade 1A formed by SARS-CoV

together with SARS-like-CoV isolated from bats and civets and clade 1B formed by SARS-CoV-2 together with SARS-like-CoV-2 isolated from bats and Malayan pangolin. Clade 2 encompasses MERS-CoV strains interspersed with camel-CoVs and also includes some bat CoV strains related with MERS-CoV. Pandemic SARS-CoV-2 strains clustered significantly close with RaTG13 CoV and pangolin CoV strains (see clade 1B). According to the CoV S protein analysis, the phylogenetic tree revealed that the distance between SARS-CoV-2 and RaTG13 CoV is shorter than that between SARS-CoV-2 and pangolin CoV. Then, pangolin CoV S protein is closest to S protein from SARS-CoV-2 behind bat RaTG13 CoV. Behind pangolin CoVs, bat CoV strains isolated in China in 2015 and 2017 are noted, followed by a cluster with bat CoVs also identified in China during 2004–2014 (clade 1B). We reinforce the scenario that suggests that SARS-CoV-2 transmission chain began from bat and reached the human. The diverse set of CoVs infecting bats frequently associated with fast-evolving CoV S protein propriety might have favored the epidemic from animal to human. The close position of SARS-like-CoV-2 isolated from pangolins indicates that this species has a potential role along the SARS-CoV-2 transmission chain (clade 1 B). Pangolin CoV strain isolated in March 2019 is very closer to RATG13 CoV and to SARS-CoV-2 than SARS-like-CoV-2 from other pangolins isolated in 2017–2018. The occurrence of recombination between SARS-like-CoVs from pangolin and bats, or even convergent evolution, must be considered as possible events for the SARS-CoV-2 origin. Phylogenetic tree showed that bat CoVs are strongly related with SARS-CoV, SARS-CoV-2, and also with MERS-CoV. Deep evolutionary and adaptation between bats and viruses contribute to asymptomatic state but spreading coronavirus to humans [25, 26]. Although bats are considered as natural hosts and reservoirs for SARS-CoV and SARS-CoV-2, camels may be the natural host of MERS-CoV [27]. The close relation between camel CoV with MERS-CoV (clade 2) indicates camel as a direct source of MERS-CoV. However, the MERS-CoV ancestor might have circulated among bat species due to MERS-like-CoV persistence in bat populations. MERS-CoV took a different evolutive way from SARS-CoV and SARS-CoV-2. In contrast to SARS-CoV and SARS-CoV-2, which use ACE2 to invade the host, MERS-CoV uses CD26 expressed by host. MERS-CoV S protein likely only recognizes conserved CD26 sequences. Furthermore, MERS-CoV S protein RBD sequence is considered more conserved comparing with SARS-CoV and SARS-CoV-2 S protein [8]. Despite the fast-evolving S protein capacity, bat CoV transmission to humans appears to be limited by a barrier. Then, intermediary hosts might be necessary to overcome genetic barriers to favor the start of human coronavirus disease outbreaks.

Host molecular repertoire used by coronavirus cycle and invasion, including ACE2 and TMPRSS2 proteins, could be a species-specific barrier to coronavirus [22]. ACE2 appears to
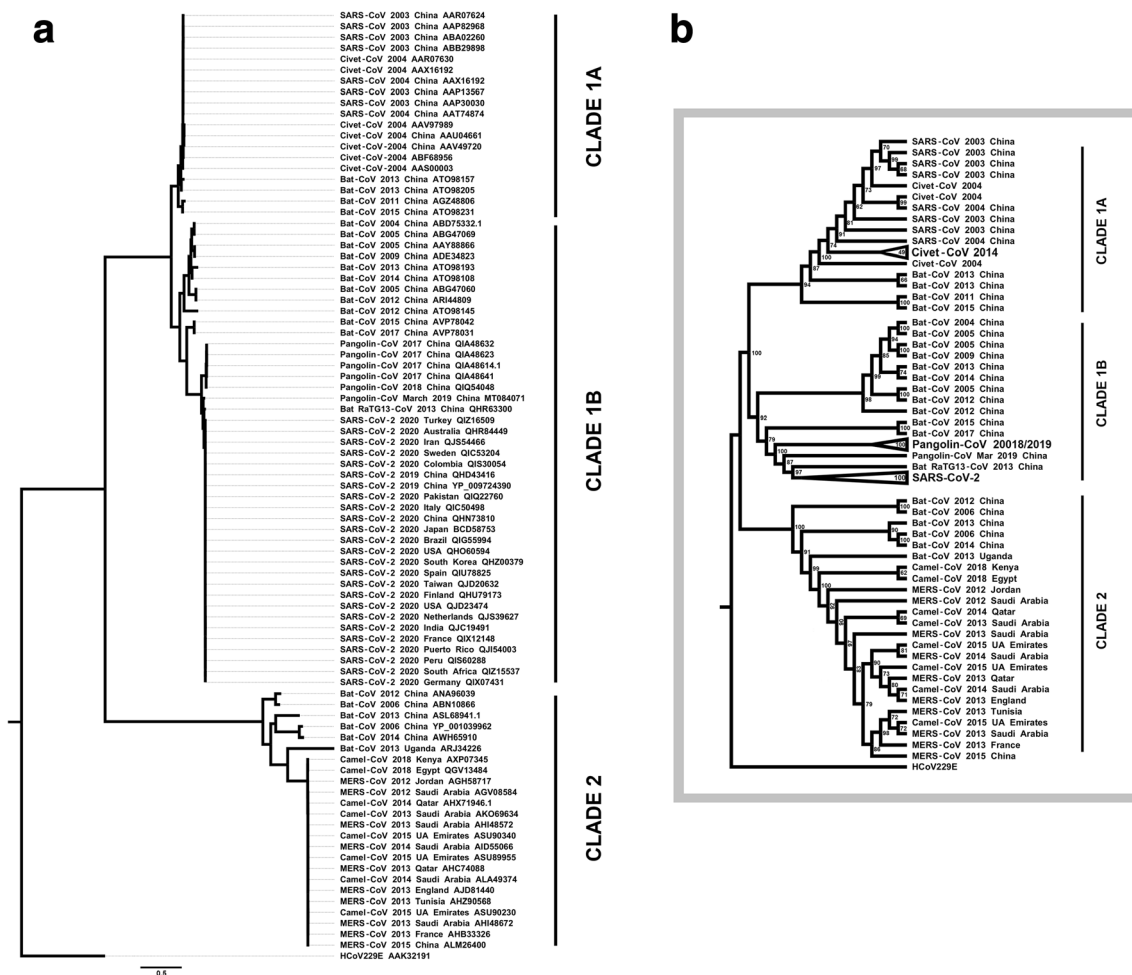
Fig. 1 CoV S protein maximum likelihood phylogenetic inference (*n* = 87). Based on the same analysis, we present (**a**) a complete phylogenetic tree and (**b**) a simplified cladogram with bootstrap values and some collapsed clades. Three clades were well clustered: clade 1A including SARS-CoV, clade 1B including SARS-CoV-2, and clade 2 including MERS-CoV, rooted by 229E CoV. Clade 1A presents SARS-CoVs interspersed with bat CoVs and civet CoVs. In the clade 1B, note that SARS-CoV-2 strain sequences are very similar to each other. SARS-CoV-2 strains were clustered near bat RaTG13 CoV and pangolin CoV strain identified in 2019. Additionally, SARS-like-CoV-2 strains identified from pangolin in 2017–2018 were also closely related to SARS-CoV-2 strains. Still in the clade 1B, there was an Asian bat SARS-like-CoV-2 cluster. Bats CoV strains are broadly distributed in tree. Clade 2 is characterized by a cluster with MERS-CoV together with camels CoV. Bats included in this cluster are more distant. Maximum likelihood phylogenetic analysis involved 87 amino acid sequences from CoV S protein

obtained from NCBI Protein database (www.ncbi.nlm.nih.gov/protein/). Malayan pangolin MT084071 sequence was translated from annotated coding region obtained from NCBI Genbank database (www.ncbi.nlm.nih.gov/genbank/). Accession codes were included in each taxon name. Phylogenetic tree inference was based in maximum likelihood method with Whelan and Goldman model [23]. The tree with the optimal log likelihood (26256.243) is shown. Bootstrap values calculated for 100 replications are shown next to the nodes of cladogram (**b**). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms. A discrete gamma distribution was used to model evolutionary rate differences among sites (4 categories [+ G, parameter = 0.7168]). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There were a total of 1494 positions in the final dataset. Phylogenetic analyses were conducted in PhYML v3.3 [24] and the trees were formatted with the FigTree v1.3.1 software (http://tree.bio.ed.ac.uk/software/figtree/).

have slow evolutionary rates between vertebrates [28]. Thus, to infer the host susceptibility, we perform a phylogenetic analysis based on 23 ACE2 amino acid sequences (Fig. 2A), including bats species, humans, primates, felids, canids, murine, and other ones. All amino acids sequences were obtained by NCBI Protein database (www.ncbi.nlm.nih.gov/protein/). ACE2 phylogenetic tree, inferred by Bayesian method, was highly supported by posterior probability values (see Fig. 2c). We also constructed a heat map based on pairwise divergence

matrix to complement phylogenetic analysis. Best substitution models to perform phylogeny and pairwise divergence matrix were selected (see Supplementary Table 1). According to our results, the phylogenetic topology conserved *Homo sapiens* closer to other primates in a monophyletic group, with low divergent amino acid sequences within this group, as expected. Bat species also composed a monophyletic group. Humans and bats are evolutionary divergent and phylogenetically distant, shown by ACE2 pairwise divergence matrix and tree,
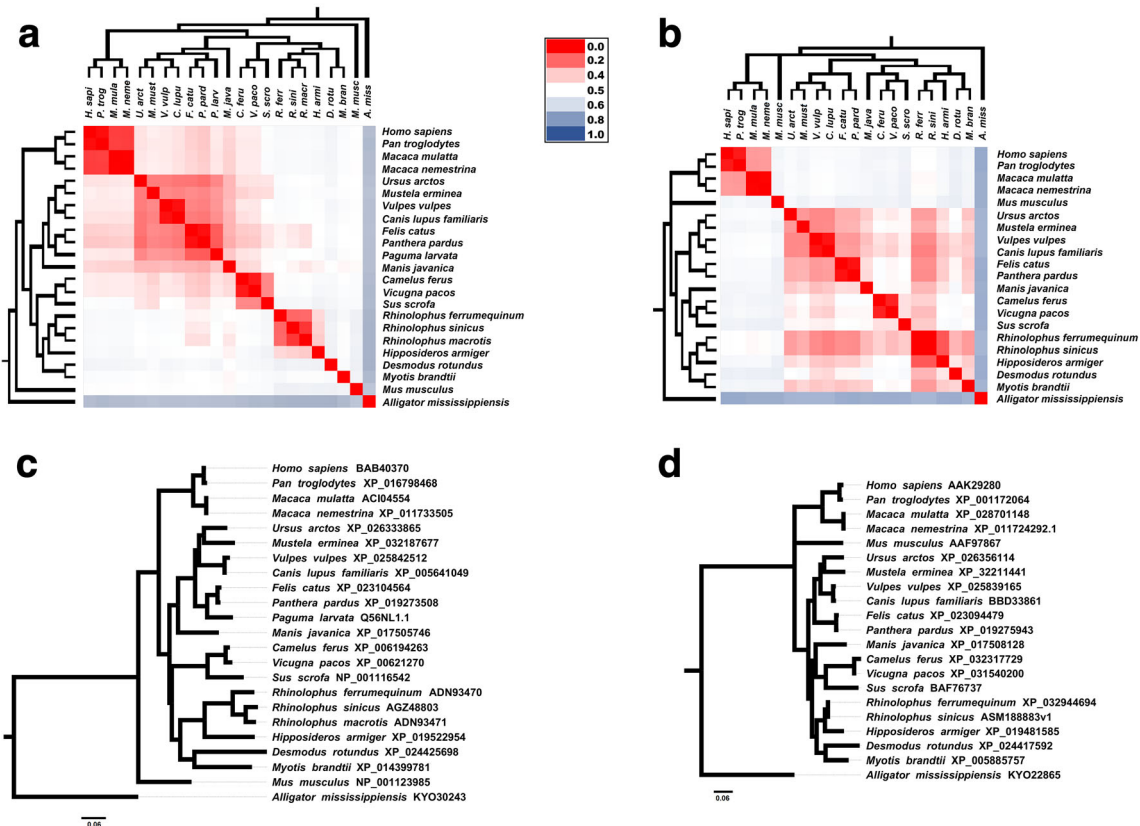
Fig. 2 Evolutionary analysis of CoV hosts: ACE2 and TMPRSS2 Bayesian phylogenetic trees associated with evolutionary divergence matrix heat maps. The heat map color gradient represents the evolutionary divergence based on the number of amino acid substitutions/site from pairwise comparison between sequences, from low (red) to high (blue). **a** ACE2-based phylogeny and heat map matrix (*n* = 23) show a cluster with primates presented low evolutionary divergences. Malayan pangolin (*Manis javanica*) and civet (*Paguma larvata*) clustered with felids, canids, and others ones besides the clade composed by bat species. Pangolin and civet are phylogenetically close from bats than with humans. On the other hand, ACE2 heat map shows that both pangolin and civet are more divergent from bats than with humans. **b** TMPRSS2 evolutionary analysis presents close relationship between pangolin and civet with bats, while primates remained distant. **c** and **d** represent more detailed Bayesian consensus phylogenetic trees based on ACE2 and TMPRSS2 with supported values described. Analysis involved 23 amino acid sequences from ACE2 protein and 21 amino acid sequences from TMPRSS2 protein. All sequences were obtained from NCBI Protein database (www.ncbi.nlm.nih.gov/protein/). *R. sinicus*

ASM188883v1 sequence was translated based on genome assembly obtained from NCBI Genome database (www.ncbi.nlm.nih.gov/genome/). (*R. macrotis* and *P. larvata* TMPRSS2 sequences were not available at the time of this analysis). Accession codes were included in each taxon name. Phylogenies were based on Bayesian analysis using JTT + G model (Jones-Taylor-Thornton model with a Gamma distribution for among-site rate variation) conducted by Mr. Bayes 3.2v [29]. Trees were searched for one million generations with sampling every 100 generations until the standard deviation from split frequencies were under 0.01. Scale bar indicates the number of substitutions/site for the trees. The parameters and the trees were summarized by wasting 25% of the samples obtained (burn-in). Phylogenetic trees were formatted with the FigTree v1.3.1 software (http://tree.bio.ed.ac.uk/software/figtree/). Evolutionary divergence between ACE2 and TMPRSS2 sequences were based on the number of amino acid substitutions per site from between sequences. Analyses were conducted using the JTT matrix-based model by MEGA X software [30], and heat maps were performed by Microsoft Excel™ software

respectively. Phylogenetic tree presented bats closer to a clade that includes pangolin and civet than humans are. Despite the phylogenetic position, humans have lower evolutionary divergence in comparison with pangolin and civet than these animals have with bats. We display the evolutionary divergence values in Table 1. Comparing humans with pangolin/civet, we found lower evolutionary divergence values with pangolin than with civet. Thus, considering ACE2 interaction as a host barrier that protects humans from zoonotic pathogens, SARS-CoV-2 in bats may require some intermediate mammalian hosts to jump to humans to start the outbreak, like civet was

required for intermediary host to favor SARS-CoV origin. We further inferred a Bayesian phylogenetic tree based on TMPRSS2 amino acid sequences (Fig. 2b), supported with high posterior probabilities values (see Fig. 2d). This tree presented pangolin close to bats. TMPRSS2 phylogeny and evolutionary divergences sequences indicated distance between *Homo sapiens* and bats/pangolin. Among all of the bat species, *Rhinolophus ferrumequinum* presented the highest similarity of TMPRSS2 amino acid sequence compared with human and pangolin. *R. ferrumequinum* is susceptible for both SARS-like-CoV and SARS-like-CoV-2 infection [10, 14].

Table 1 Estimates of evolutionary divergence between ACE2 protein sequences

Species	H. sap	P. tro	P. lar	M. jav	R. fer	R. sin	R. mac	H. arm	D. rot	M. bra	M. mus
Homo sapiens											
Pan troglodytes	0.01026										
Paguma larvata	**0.22463**	0.22472									
Manis javanica	**0.20912**	0.20570	0.18750								
Rhinolophus ferrumequinum	0.28301	0.28007	**0.26965**	**0.25506**							
Rhinolophus sinicus	0.27941	0.27520	**0.25946**	**0.24230**	0.09058						
Rhinolophus macrotis	0.27768	0.27505	**0.25358**	**0.22599**	0.09273	0.04557					
Hipposideros armiger	0.28345	0.27822	**0.26322**	**0.23590**	0.20608	0.20141	0.19348				
Desmodus rotundus	0.31272	0.30576	**0.31213**	**0.26304**	0.34012	0.32423	0.32713	0.33562			
Myotis brandtii	0.26953	0.26678	**0.26588**	**0.24297**	0.29533	0.26611	0.26925	0.26943	0.27961		
Mus musculus	0.25338	0.25776	0.26082	0.24598	0.32786	0.32524	0.31852	0.33391	0.33965	0.31988	

The numbers of amino acid substitutions per site between sequences are shown. Divergence values involving civet and/or pangolin (possible intermediary hosts) with human and/or bats were presented in bold. Pangolin and civet have lower evolutionary distance from humans than from bats. Evolutionary analyses were made using the JTT matrix-based model [31], conducted in MEGA X software [30]. The rate variation among sites was modeled with a gamma distribution (shape parameter = 0.46). This analysis involved 11 amino acid sequences aligned by MUSCLE [32]. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 829 positions in the final dataset

Molecular similarity of TMPRSS2 from SARS-CoV and SARS-CoV-2 susceptible hosts is suggestive that the molecule seems not to be an important barrier to emerging human coronaviruses. Probably, the SARS-CoV-2 entry into host cells, interaction with ACE2 receptor, appears to be the main challenge. While most of SARS-CoV-2 phylogenetic studies are focusing in viral phylogenies, we contributed with host analysis based on ACE2 and TMPRSS2. In summary, we report and conclude that SARS-like-CoV-2 strains infected pangolin and bats and are phylogenetically close to SARS-CoV-2. Thus, our phylogenetic analysis based on SARS-CoV-2 S protein supports the hypothesis of SARS-CoV-2 transmission chain began from bat, had Malayan pangolin as intermediary host, and infected humans. Considering CoV S protein as the key to invade host cell through host ACE2, we explore this host molecule as base of analysis to understand the role of the hosts in the origin of SARS-CoV-2. ACE2 sequence from pangolin has low evolutionary divergence compared with humans but is more divergent from bats. Taken together, combined viral and host evolutive analysis corroborated the hypothesis of Malayan pangolin as intermediary host in SARS-CoV-2 origin. Looking back through coronavirus outbreak histories, wild animal chains appear to be necessary. Frequent jumps of bats virus produce potential infections or short transmission chains that resolve, with no adaptation to sustained transmission [33].

Therefore, repeated opportunities may promote zoonotic events resulting in coronavirus outbreaks. For SARS-CoV, there were broad evidences suggesting civet as the main intermediary host [7, 34, 35]. Our results also indicated civet as an important player in SARS-CoV origin. Civet might be infected by SARS-like-CoV and has intermediary ACE2 divergence between humans and bats. Comparing SARS-CoV with SARS-CoV-2 origins, pangolin ACE2 amino acid sequence has yet lower evolutionary divergence with humans, while civet ACE2 sequence is more divergent compared with humans. Thus, pangolin has become an opportune host to intermediates bat-to-human SARS-CoV-2 jump and entry. Differently from bats, which are able to suppress viral replication, pangolin is an amplifying host which allows the increase of viral load and probably favored SARS-CoV-2 jump to human host and human-to-human transmission subsequently. The recurrent emergence of zoonotic disease outbreaks caused by coronavirus alerts once more for the implementation of strict rules to decrease or eliminate consumption and domestication or even the ban on wildlife markets.

Compliance with ethical standards

References

1. Zhou P, Yang XL, Wang XG et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579(7798):270–273. https://doi.org/10.1038/s41586-020-2012-7

2. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, Tan KS, Wang DY, Yan Y (2020) The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- an update on the status. Mil Med Res 7:11. https://doi.org/10.1186/s40779-020-00240-0

3. Drosten C (2013) Is MERS another SARS? Lancet Infect Dis 13(9):727–728. https://doi.org/10.1016/S1473-3099(13)70159-2

4. Masters PS (2006) The molecular biology of coronaviruses. Adv Virus Res 65:193–292. https://doi.org/10.1016/S0065-3527(06)66005-3

5. Agnihothram S, Yount BL, Donaldson EF et al (2014) A mouse model for Betacoronavirus subgroup 2c using a bat coronavirus strain HKU5 variant. MBio. 5:e00047–e00014. https://doi.org/10.1128/mBio.00047-14

6. Li W, Wong S-K, Li F, Kuhn JH, Huang IC, Choe H, Farzan M (2006) Animal origins of the severe acute respiratory syndrome coronavirus:iInsight from ACE2-S-protein interactions. J Virol 80(9):4211–4219. https://doi.org/10.1128/jvi.80.9.4211-4219.2006

7. Wang LF, Eaton BT (2007) Bats, civets and the emergence of SARS. Curr Top Microbiol Immunol 315:325–344. https://doi.org/10.1007/978-3-540-70962-6_13

8. Lu G, Wang Q, Gao GF (2015) Bat-to-human: spike features determining "host jump" of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiol 23(8):468–478. https://doi.org/10.1016/j.tim.2015.06.003

9. Sun J, He WT, Wang L, Lai A, Ji X, Zhai X, Li G, Suchard MA, Tian J, Zhou J, Veit M, Su S (2020) COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends Mol Med 26(5):483–495. https://doi.org/10.1016/j.molmed.2020.02.008

10. Lam TTY, Shum MHH, Zhu HC et al (2020) Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature. https://doi.org/10.1038/s41586-020-2169-0

11. Malik YS, Sircar S, Bhat S, Sharun K, Dhama K, Dadar M, Tiwari R, Chaicumpa W (2020) Emerging novel coronavirus (2019-nCoV)—current scenario, evolutionary perspective based on genome analysis and recent developments. Vet Q 40(1):68–76. https://doi.org/10.1080/01652176.2020.1727993

12. Cohen J (2020) Wuhan seafood market may not be source of novel virus spreading globally. Science (80-). https://doi.org/10.1126/science.abb0611

13. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223):497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

14. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang LF (2005) Bats are natural reservoirs of SARS-like coronaviruses. Science 310(5748):676–679. https://doi.org/10.1126/science.1118391

15. Dobson AP (2005) What links bats to emerging infectious diseases? Science 310(5748):628–629. https://doi.org/10.1126/science.1120872

16. Rothan HA, Byrareddy SN (2020) The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 109:102433. https://doi.org/10.1016/j.jaut.2020.102433

17. Ding Y, He L, Zhang Q, Huang Z, Che X, Hou J, Wang H, Shen H, Qiu L, Li Z, Geng J, Cai J, Han H, Li X, Kang W, Weng D, Liang P, Jiang S (2004) Organ distribution of severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV) in SARS patients: implications for pathogenesis virus transmission pathways. J Pathol 203(2):622–630. https://doi.org/10.1002/path.1560

18. Jia HP, Look DC, Shi L, Hickey M, Pewe L, Netland J, Farzan M, Wohlford-Lenane C, Perlman S, McCray PB Jr (2005) ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. J Virol 79(23):14614–14621. https://doi.org/10.1128/jvi.79.23.14614-14621.2005

19. Hoffmann M, Kleine-Weber H, Schroeder S et al (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181(2):271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052

20. Glowacka I, Bertram S, Muller MA, Allen P, Soilleux E, Pfefferle S, Steffen I, Tsegaye TS, He Y, Gnirss K, Niemeyer D, Schneider H, Drosten C, Pohlmann S (2011) Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. J Virol 85(9):4122–4134. https://doi.org/10.1128/jvi.02232-10

21. Iwata-Yoshikawa N, Okamura T, Shimizu Y, Hasegawa H, Takeda M, Nagata N (2019) TMPRSS2 contributes to virus spread and immunopathology in the airways of murine models after coronavirus infection. J Virol 93. https://doi.org/10.1128/jvi.01815-18

22. Li F (2013) Receptor recognition and cross-species infections of SARS coronavirus. Antivir Res 100(1):246–254. https://doi.org/10.1016/j.antiviral.2013.08.014

23. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18(5):691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851

24. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59(3):307–321. https://doi.org/10.1093/sysbio/syq010

25. Balboni A, Battilani M, Prosperi S (2012) The SARS-like coronaviruses: the role of bats and evolutionary relationships with SARS coronavirus. New Microbiol 35(1):1–16

26. Brook CE, Dobson AP (2015) Bats as "special" reservoirs for emerging zoonotic pathogens. Trends Microbiol 23(3):172–180. https://doi.org/10.1016/j.tim.2014.12.004

27. Hayman DTS (2016) Bats as viral reservoirs. Annual Review of Virology 3(1):77–99. https://doi.org/10.1146/annurev-virology-110615-042203

28. Lv Y, Li Y, Yi Y, Zhang L, Shi Q, Yang J (2018) A genomic survey of angiotensin-converting enzymes provides novel insights into their molecular evolution in vertebrates. Molecules. 23. https://doi.org/10.3390/molecules23112923

29. Ronquist F, Teslenko M, Van Der Mark P et al (2012) Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61(3):539–542. https://doi.org/10.1093/sysbio/sys029

30. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6):1547–1549. https://doi.org/10.1093/molbev

31. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8(3):275–282. https://doi.org/10.1093/bioinformatics/8.3.275

32. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797. https://doi.org/10.1093/nar/gkh340

33. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF (2020) The proximal origin of SARS-CoV-2. Nat Med 26:450–452. https://doi.org/10.1038/s41591-020-0820-9

34. Song HD, Tu CC, Zhang GW, Wang SY, Zheng K, Lei LC, Chen QX, Gao YW, Zhou HQ, Xiang H, Zheng HJ, Chern SWW, Cheng F, Pan CM, Xuan H, Chen SJ, Luo HM, Zhou DH, Liu YF, He JF, Qin PZ, Li LH, Ren YQ, Liang WJ, Yu YD, Anderson L, Wang M, Xu RH, Wu XW, Zheng HY, Chen JD, Liang G, Gao Y, Liao M, Fang L, Jiang LY, Li H, Chen F, di B, He LJ, Lin JY, Tong S, Kong X, du L, Hao P, Tang H, Bernini A, Yu XJ, Spiga O, Guo ZM, Pan HY, He WZ, Manuguerra JC, Fontanet A, Danchin A, Niccolai N, Li YX, Wu CI, Zhao GP (2005) Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc Natl Acad Sci U S A 102(7):2430–2435. https://doi.org/10.1073/pnas.0409608102

35. Guan Y, Zheng BJ, He YQ et al (2003) Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302(5643):276–278. https://doi.org/10.1126/science.1087139

36. Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.