



Origin and Evolution of SARS-CoV-2: An Enigma

Nafiz M, Islam KU, Bhat ZA and Iqbal J*

Multidisciplinary Center for Advanced Research and Studies, Jamia Millia Islamia, New Delhi, India

*Corresponding author: Jawed Iqbal, Multidisciplinary Center for Advanced Research and Studies, Jamia Millia Islamia, India, Tel: 9911053902; Email: jiqbal1@jmi.ac.in

Review Article

Volume 8 Issue 1

Received Date: January 02, 2024

Published Date: February 02, 2024

DOI: 10.23880/vij-16000338

Abstract

SARS-CoV-2 has caught much attention due to its highly contagious nature which caused COVID-19 outbreak. Viruses have been known to evolve in response to natural selection and as a result of drug-based interventions mainly by incorporating mutations. The substantial effect of SARS-CoV-2 variants on its transmission and disease severity is prompting scientists to explore the evolutionary processes. This review summarizes different theories of SARS-CoV-2 evolution that have been put forward by different scientific schools of thought to describe the origin of SARS-CoV-2 and the possibility of bats for being its natural reservoirs. Furthermore, insights from the phylogenetic tree analysis depict divergence of VOCs of SARS-CoV-2 from the original strain in a short span of time. As evident from the genomic and phylogenetic studies, bats host a range of coronaviruses having ability for inter-species host switching, there is high probability for the potential spillovers that might lead to COVID-19-like pandemics in future.

Keywords: SARS-CoV-2; COVID-19; Zoonotic Transmission; Evolution; Variants of Concern

Introduction

Coronaviruses belong to the family *Coronaviridae*, and on the basis of the genomic structure and phylogenetic relationship, this family is further classified into different genera - α -, β -, γ - and δ - coronavirus [1]. Coronaviruses had gained minimal research focus until 2003, after which the outbreak of SARS (severe acute respiratory syndrome) caused by the SARS-CoV renewed its research attention altogether and brought this family of viruses into the limelight of research and scientific interventions [2]. Ten years later, a major outbreak of MERS (Middle East Respiratory Syndrome), followed by the recent emergence of SARS-CoV-2, one of the deadliest coronaviruses so far has made it imperative to understand the origin and evolution of these viruses. Both SARS and MERS originated from bats and the zoonotic origin of these viruses allows them to jump from one host to another. Virologists had warned the presence of a large reservoir of SARS-CoV-like viruses in horseshoe bats (*Rhinolophus*), thus underscoring the significance of finding evolutionary roots of these viruses [3]. Amid the global

public health emergency of COVID-19, it is quite reasonable to explore the origins of the pandemic. The initial infection of humans by SARS-CoV-2 can be delineated by two probable scenarios- zoonotic transfer from wild animals and a leak of the pathogen from a research laboratory. Thus, there is a pressing need to understand in detail how an animal virus can jump between species in order to document the intermediate host of the pathogen. Although the zoonotic transfer is supported by stronger evidence, the laboratory leak hypothesis is still debated. Thus, reaching a factual conclusion is of great essence in order to prevent any future pandemic.

Substantiation for the Zoonotic Origin of SARS-CoV-2

Coronaviruses (CoVs) can infect animals including humans and can cause a spectrum of health consequences ranging from mild disease to death [4]. Humans have encountered several CoVs till date and SARS-CoV-2 is the ninth

documented CoV that infect humans [5]. Table 1 summarizes different hCoVs and their respective hosts and intermediate hosts. All the CoVs including SARS-CoV-1 and MERS that had a previous encounter with humans had a zoonotic origin as reviewed by Holmes, et al. [6]. Sequence analysis of the original strain of SARS-CoV-2 during the early stages of pandemic revealed that this virus has a close similarity (88%) with two bat-derived CoVs: SL-CoVZC45 and SL-CoVZXC21, emphasizing the perception that bats could be the original reservoirs of CoVs [7]. Numerous studies have revealed that

the horseshoe bats in China harbor coronaviruses, collectively referred to as SARS-related coronavirus (SARSr-CoV) which belong to Sarbecovirus subgenus of the Betacoronavirus genus [8]. SARS-CoV-2 shares a 96.2% nucleotide sequence identity with the bat coronavirus -RaTG13 isolated from horseshoe bat *Rhinolophus affinis*. It was later concluded that the progenitor of SARS-CoV was seeded in an intermediate animal host by a Sarbecovirus harboring in horseshoe bats [9].

Human CoVs	Classification	Linage	Natural Host	Intermediate Host	Reference
HCoV-OC43	Beta- CoV	A	Rodents	Bovines	[1]
HCoV-229E	Apha-CoV	-	Bats	Camelids	[2]
HCoV-NL63	Alpha-CoV	-	Bats	-	[1]
HCoV-HKU1	Beta -CoV	A	Rodents	-	[2]
MERS-CoV	Beta-CoV	C		Dromedary Camels	[3]
SARS-CoV-1	Beta-CoV	B	Bats	Palm Civets	[3]
SARS-CoV-2	Beta-CoV	B	Bats		[4]

Table 1: Different Human Coronaviruses and their hosts/intermediate host.

The emergence of SARS-CoV-2 in China led to an enigma that how did SARS-CoV-2 reach Wuhan city, located >1500 km from Yunnan province - the closest known location from where horseshoe bats harboring Sarbecovirus have been collected. This initial strain was detected by sequencing of viral genome, which also paved way for the detection of other emerging strains by laying the basis for the generation of phylogenetic tree, as reviewed by Islam and Iqbal [10]. By extensive sampling, it was found that horseshoe bats harboring coronaviruses which are genetically close to SARS-CoV-2 are dispersed widely from East to West China, Southeast Asia and Japan [9]. Owing to the broad geographic distribution of these probable reservoir hosts, it is misleading to specifically take Yunnan into focus. Sequence analysis revealed that the evolutionarily closest bat Sarbecoviruses share a common ancestor with SARS-CoV-2 at least 40 years ago, implying that the Yunnan-collected viruses are highly divergent from the SARS-CoV-2 progenitor [9]. Of these, the first Sarbecovirus reported by Wuhan Institute of Virology (WIV), RaTG13 is highly divergent and could not have been the SARS-CoV-2 progenitor, accentuating critical evidence against “lab-leak” theory.

The animal origin of SARS-CoV-2, whether directly or via an intermediate host, was first speculated by the Chinese virology research team of the Wuhan laboratory. They had the notion that origin of COVID-19 is associated with wet markets of Wuhan. Although the animal source of SARS-CoV-2 is not certain, but it is noteworthy that live wild animals

susceptible to Sarbecoviruses, like minks, foxes, civet cats, raccoon dogs, were being sold there, in particular in Huanan market (considered as the epicenter of the outbreak). There is no origin or quarantine certificate guarantying the hygiene of live animals sold there for food or as pets. On-site slaughtering of animals with arguable hygienic practices poses a high risk of infection. Furthermore, the respiratory aerosols from infected animals could be a possible source of infection. Investigation of Wuhan animal market by WHO team suggested that pangolins (Pholidata) can be considered as a plausible intermediate host of SARS-CoV-2. But, just like with bats, viruses isolated from pangolins are evolutionarily quite distant from SARS-CoV-2 [11].

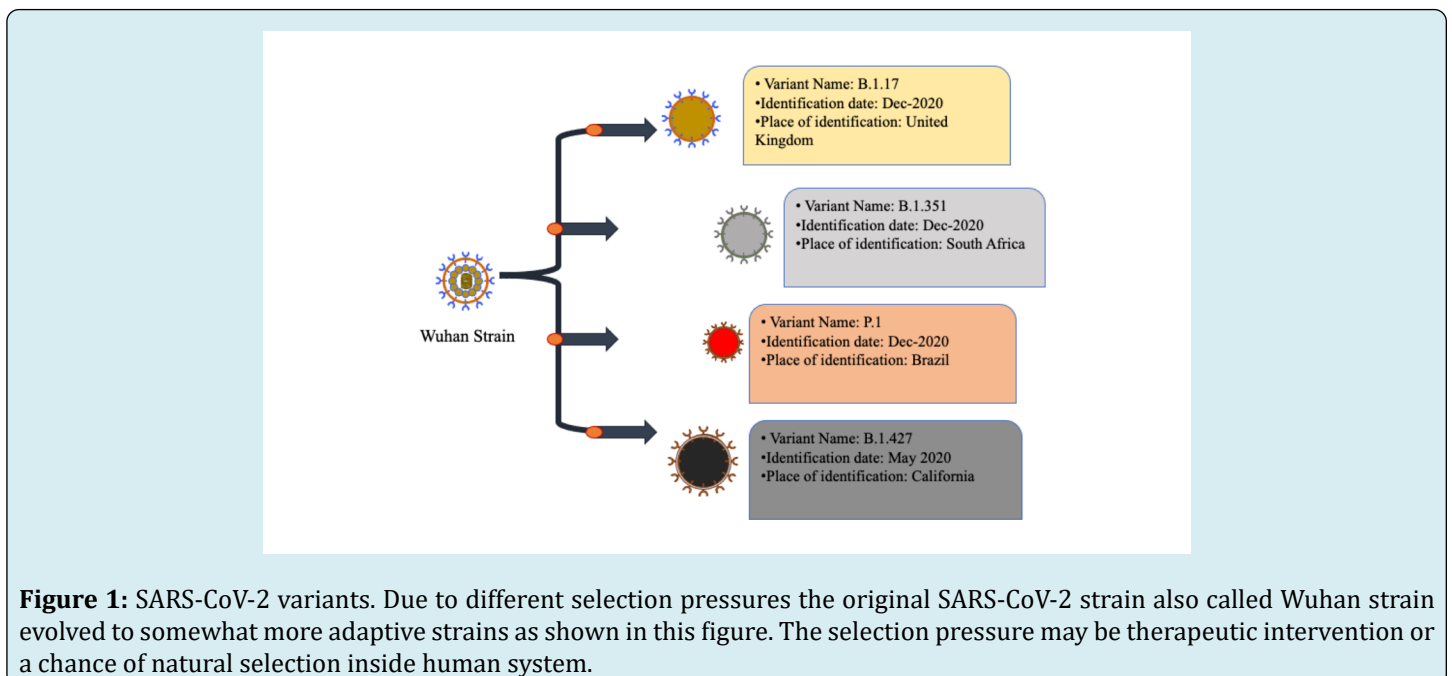
A plethora of evidence authenticates the SARS-CoV-2 emergence as a natural spillover event [12]. Certain studies reveal that Chinese animal traders in 2003 had high serum IgGs against SARS-CoV-1 emphasizing on the regular exposure of this virus from the wet animal market which further provides evidence of human exposure to these viruses. Albeit the transfer of Sarbecovirus having genome closely related to SARS-CoV-2 from a bat cave to humans is possible, the event is extremely rare relative to the frequency of animal contacts occurring in animal trading. On the other hand, bat caves are being visited more often to collect bat fecal matter for use as fertilizer rather than for research purposes, consistent with the finding of ongoing SARSr-CoV transmissions to humans in rural areas [13].

Phylogeny and Emergence of SARS-CoV-2 Variants of Concern

Coronaviruses found to be genetically diverse yet share the common ancestry as evident from the genomic and phylogenetic analysis carried out over the decades. Belonging to the *Coronaviridae* family, it comprises of multiple genera including *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* based on their phylogeny [2]. Years of research has led to the conclusion that Alpha and Beta coronaviruses can potentially infect mammals only whereas Gamma and Delta coronaviruses mostly infect birds [14]. Studied in detail and most well-known to the humankind, SARS-CoV and MERS are the two representatives of *Betacoronaviruses* that led to the earliest pandemics of 21st century wherein mainly the respiratory diseases (severe acute respiratory syndrome– SARS) caused the death of hundreds of people [15]. These coronaviruses are found to have potentially jumped from species to species wherein bats are thought to be the major reservoirs for most of the coronaviruses and multiple mammalian species (Civets and Camels in case of SARS – CoV and MERS respectively) can act as the intermediates before infecting the humans [16]. It is well known that if a virus is evolving inside the host, then it is quite expected that it would be adapting to enter the cells efficiently and evade the immune response, thereby giving rise to a variant with augmented transmissibility and better immune evasion. Coronavirus evolution is driven by several factors which include their genetic diversity, larger genome size, high rate of mutations owing to the low fidelity polymerases [17]. Since the COVID-19 pandemic is still going on, the world is thus witnessing the rise of

different variants of SARS-CoV-2 in different parts of the globe. As evidenced from the study by Singh et al, there is an increasing trend of positively selected mutations in receptor binding domain (RBD) of several SARS-CoVs, consequently making the virus prone to evolve its interaction with the host [18]. CoVs undergo positive selection in the residues within RBD, facilitating efficient interactions with host cellular proteins. Thus, SARS-CoV-2 RBD is rapidly evolving, leading to the generation of novel variants which are associated with greater transmissibility, altered virulence, and the ability to evade natural and vaccine-mediated immunity. Such variants are collectively referred to as Variants of Concern (VOC) [18].

Genomic mutations give rise to new variants and the closely related variants of a particular strain are grouped together in clades in a phylogenetic tree. There are nine clades of SARS-CoV-2 variants [18]. Highly transmissible variant B.1.1.7 (referred to as Alpha variant) that originated in United Kingdom has been observed to carry non-synonymous mutations or deletions as shown in Figure 1. Similarly, B.1.617.2 variant (Delta) was identified firstly in India which possess the positively selected mutations in its spike protein [19]. All these represent SARS-CoV-2 variants of concern (VoC) currently circulating in the human population. The recent variant B.1.1.529 (Omicron) is the most profoundly mutated variant among all the VOCs identified so far, having enhanced transmissibility and partial resistance to vaccine-induced immunity [20]. Few of such variants have been observed to possess a potential to escape vaccine mediated immunity thus, a sign of concern as vaccine mediated immunity can impart selective pressure leading to SARS-CoV-2 evolution.



The occurrences of spillover have been witnessed in the past and Swine Acute Diarrhoea Virus (SADS-CoV) being the recent one from bats to pigs [20]. COVID-19 affecting almost entire world with about 450 million infections and more than 6 million deaths (<https://www.worldometers.info/coronavirus/>), being the most recent pandemic of 21st century, might be yet another spillover as evident from the phylogenetic tree analysis (Figure 2). It was found to be most closely related to bat coronaviruses including RpYN06 and RaTG13 as compared to SARS-CoV and MERS and has been found to share common ancestry with multiple bat coronaviruses. In turn, SARS-CoV and MERS also share common ancestry with SARS-CoV-2. Hence, from the phylogenetic tree analysis, significant evidence points towards bat for being the natural reservoir for coronaviruses which may potentially jump from one species to another due

to the incorporation of mutations. Another interesting finding from this analysis is the divergence of multiple variants of SARS-CoV-2 (called Variants of Concern (VOC)) wherein Omicron and Beta variants seem to relatively diverge from the original strain and their distinctive phenotypic characters are poorly explored so far. The probable explanation may be that virus is trying to adapt and get stabilized in the host (human beings in this case) as has been found in the case of MERS wherein it is said to have jumped from bat to camels and stayed there for almost 30 years before jumping again to another host (human beings) [16]. Since, bats host vast number of coronaviruses with the property of inter-species host switching, there would always be chances for the viral spillovers resulting in pandemics of COVID-19 scale or even worse.

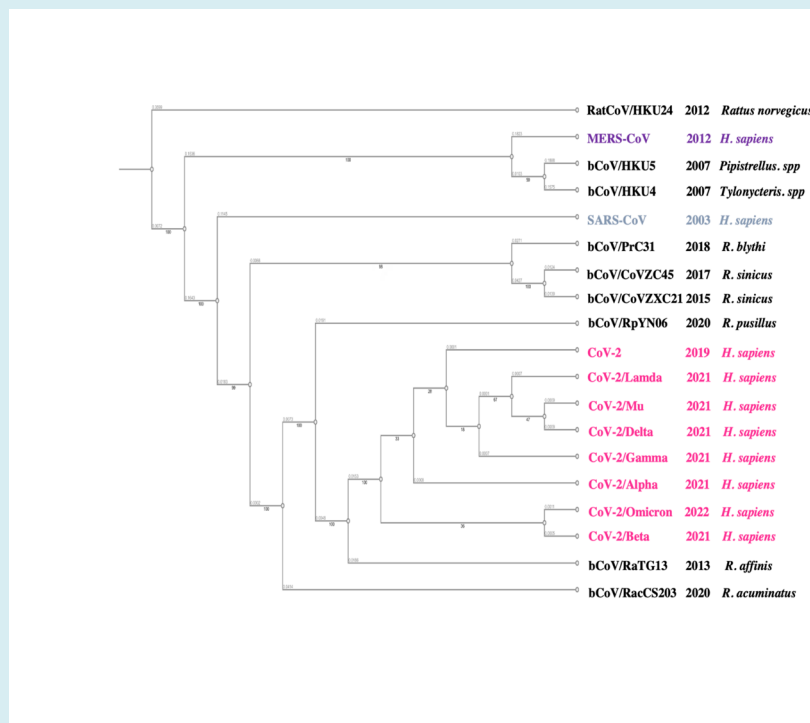


Figure 2: Phylogenetic tree showing common ancestry between different coronaviruses. All the coronaviruses have been shown to share the common ancestry, though SARS-CoV-2 may be more closely related to bat coronavirus RpYN06 and RaTG13. Even among the multiple variants of concern (VOC) in SARS-CoV-2, Omicron and Beta variant seems to diverge from the original variant whereas, Gamma, Delta, Lambda and Mu are more closely related to the original strain.

Natural Selection in the Origin of SARS-CoV-2

The RBD of SARS-CoV-2 has evolved aptly to bind efficiently to the human angiotensin converting enzyme 2 (ACE2) receptor, which constitutes the primary binding site for the virus entry into the host cells. There is significant evidence vindicating the theory that the virus has not been

generated by genetic manipulations [2,21,22]. The origin of SARS-CoV-2 can be explained in terms of natural selection by two probable scenarios - before and after the zoonotic transfer of the virus.

Researchers have modeled the binding affinity of Spike receptor binding domain and host ACE2 with the aim to identify the potential intermediate host of the virus [23].

It is contemplated that the SARS-CoV-2 was transmitted to humans via some intermediate host, during which the progenitor procured the critical residues conferring ACE2 binding ability along with furin cleavage site. This is partly supported by the fact that the ACE2-binding residues of CoV isolated from pangolin (*Manis javanica*) are almost identical to that of SARS-CoV-2 [24]. Phylogenetic analyses revealed that some pangolin-CoVs are genetically similar to SARS-CoV-2 but do not significantly support the emergence of SARS-CoV-2 directly from these pangolin-CoVs [25]. Furthermore, furin-cleavage site is neither found in bat-CoVs nor in pangolin-CoVs. Mutations like indels in the CoV genome can generate furin cleavage site, implying that cleavage site can arise by an evolutionary process [21]. However, for the natural selection to operate efficiently, it is imperative to have high population density of the progenitor virus, which is possible in natural milieu and not in laboratory setup.

It is also probable that the natural selection could have occurred after the zoonotic transfer, when the progenitor virus had gained entry into the human host and propagated there during undetected human-to-human transmissions [26] a large number of SARS-related coronaviruses (SARSr-CoVs). Once the virus acquired the genomic features critical for its pathogenicity, the pandemic was set off.

The “Lab-Leak” Hypothesis

Although there are strong evidences supporting the zoonotic origin of the SARS-CoV-2, the lab leak theory has not been obviated. Some researchers have a vocation for thorough investigations into the hypothesis that the virus originated from Wuhan Institute of Virology (WIV), which has been carrying out research on bat coronaviruses from past several years. It is a BSL-4 laboratory which authorizes the institute to work with the most hazardous biological agents. The laboratory is known to generate mutations in the CoV genomes and one such study involved genetic manipulation in the Spike protein of virus. This mutation granted the viruses ability to bind to ACE2 receptor and hence infect the human cells, which otherwise it was unable to do. Another argument suggests that the virus bears atypical genetic features, hinting that it was created in lab. The presence of a rare codon of Arginine (CGG) has been speculated as sign that the virus was presumably generated in the lab [11]. Two out of 42 arginine in the spike protein are encoded by these rare CGG tandem codons and that too are present in the furin cleavage site. Curiously, these doublet CGG-CGG codons are present in SARS-CoV-2 only and not in the furin cleavage sites of any other related CoVs [27].

Despite these assertions and contentions, it is still believed that the virus could not have been engineered. Just because the virus can dwell and spread among humans doesn't

justify the theory that it was designed to do so. The virus can propagate well among minks and other animals as well. Besides, the virus transmission in humans wasn't efficient enough [28]. This is supported by the fact that new and more competent variants have evolved globally. For example, the Delta variant (B.1.617.2) – a highly transmissible variant of SARS-CoV-2 first reported in India bearing mutations in the furin cleavage site, making the virus more infective.

Although there is ample evidence in support of the zoonotic origin of the virus, there are minor ambiguities which pave way for the laboratory escape theory. Theoretically, it is quite possible that CoVs gained mutations in Spike proteins during cell culture passages at WIV. Nonetheless, this would've required the prior isolation of progenitor virus having high genetic similarity, which however hasn't been still identified [2]. Such uncertainties need to be withdrawn in order to reach a significant conclusion regarding the origin of SARS-CoV-2 because a clear knowledge can assist in preventing any such pandemic in future.

Conclusion and Future Perspectives

Amidst the global pandemic of COVID-19, it is of great essence to ponder on the question that how SARS-CoV-2 originated. Based on the scientific evidences gathered so far, the most plausible justification for the origin of SARS-CoV-2 is the zoonotic event. The epidemiology of the virus is somewhat related to the previous animal-market associated CoV outbreaks. Thorough comprehension of the mechanism involved in the jumping of virus between species boundaries can help to avert such zoonotic events in future. Molecular and phylogenetic analyses provide an indication that SARS-CoV-2 has been originated from a non-human virus reservoir. Many theories regarding SARS-CoV-2 origin have been put forward by different scientific school of thoughts with their own pros and cons. There is not enough information available to claim the origin and species jumping of SARS-CoV-2, however, data related to this is constantly updating. Although the likelihood of a laboratory leak cannot be entirely set aside, and is dismissed, and may be near impossible to falsify, this conduit for emergence is highly unlikely relative to the numerous and repeated human-animal contacts that occur routinely in the wildlife trade. Some research groups claim the laboratory escape of SARS-CoV-2 but there is no substantial documented evidence regarding their claim.

COVID-19 pandemic ramped up the data generation and data sharing in the public domain enabling scientists to trace the evolutionary track in real time (<https://www.ncbi.nlm.nih.gov/sars-cov-2/>) but there are still many knowledge gaps that needed to be filled. In order to finally reach to certain outcome, a transparency is needed in sharing the

data related to the evolutionary tracing of SARS-CoV-2.

Acknowledgement

JI acknowledges Ramalingaswami Fellowship Grant (BT/RLF/Re-entry/09/2015) from Department of Biotechnology (DBT) and Early Career Research Award Grant (File No. ECR/2018/002114) from the Science and Engineering Research Board (SERB), Department of Science and Technology, Govt. of India.

Funding

There is no funding involved in this article.

Author Contributions

MN, KU and JI conceptualized the idea. KU and MN wrote the major portion of the manuscript and JI modified and corrected. ZAB created phylogenetic tree. Authors finalized the article and approved the final version for submission.

Competing Interest

The authors declare no conflict of interest.

Research Involving Human Participants or Animals

This article does not contain any studies with animals or humans performed by any of the authors

Informed Consent

This article does not require informed consent as no human and animal data has been collected in this study.

References

1. Corman VM, Muth D, Niemeyer D, Drosten C (2018) Hosts and Sources of Endemic Human Coronaviruses. *Advances in Virus Research* 100: 163-188.
2. Cui J, Li F, Shi ZL (2019) Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 17: 181-192.
3. Morens DM, Breman JG, Calisher CH, Doherty PC, Hahn BH, et al. (2020) The Origin of COVID-19 and Why It Matters. *The American Journal of Tropical Medicine and Hygiene* 103(3): 955-959.
4. Su S, Wong G, Shi W, Liu J, Lai ACK, et al. (2016) Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends in Microbiology* 24(6): 490-502.
5. Lednicky JA, Tagliamonte MS, White SK, Elbadry MA, Alam MM, et al. (2021) Emergence of porcine delta-coronavirus pathogenic infections among children in Haiti through independent zoonoses and convergent evolution. medRxiv.
6. Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, et al. (2021) The origins of SARS-CoV-2: A critical review. *Cell* 184(19): 4848-4856.
7. Lu R, Zhao X, Li J, Niu P, Yang B, et al. (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395(10224): 565-574.
8. Li W, Shi Z, Yu M, Ren W, Smith C, et al. (2005) Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science* 310(5748): 676-679.
9. Lytras S, Xia W, Hughes J, Jiang X, Robertson DL (2021) The animal origin of SARS-CoV-2. *Science* 373(6558): 968-970.
10. Islam KU, Iqbal J (2020) An Update on Molecular Diagnostics for COVID-19. *Front Cell Infect Microbiol* 10: 560616.
11. Ruiz-Medina BE, Varela-Ramirez A, Kirken RA, Robles-Escajeda E (2022) The SARS-CoV-2 origin dilemma: Zoonotic transfer or laboratory leak?. *BioEssays* 44(1): 2100189.
12. Lytras S, Hughes J, Martin D, Swanepoel P, de Klerket A, et al. (2022) Exploring the Natural Origins of SARS-CoV-2 in the Light of Recombination. *Genome Biology and Evolution* 14(2): evac018.
13. Li H, Mendelsohn E, Zong C, Zhang W, Hagan E, et al. (2019) Human-animal interactions and bat coronavirus spillover potential among rural residents in Southern China. *Biosafety and Health* 1(2): 84-90.
14. Woo PCY, Lau SKP, Lam CSF, Lau CCY, Tsanget AKL, et al. (2012) Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus. *J Virol* 86(7): 3995-4008.
15. Abdelrahman Z, Li M, Wang X (2020) Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses. *Front Immunol* 11: 552909.

16. Meyer B, Drosten C, Müller MA (2014) Serological assays for emerging coronaviruses: Challenges and pitfalls. *Virus Research* 194: 175-183.
17. Vijgen L, Lemey P, Keyaerts E, Van Ranst M (2005) Genetic Variability of Human Respiratory Coronavirus OC43. *J Virol* 79(5): 3223-3225.
18. Singh J, Pandit P, McArthur AG, Banerjee A, Mossman K (2021) Evolutionary trajectory of SARS-CoV-2 and emerging variants. *Virol J* 18: 166.
19. Cherian S, Potdar V, Jadhav S, Yadav P, Gupta N, et al. (2021) Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. *Molecular Biology, biorxiv*.
20. Araf Y, Akter F, Tang YD, Fatemi R, Parvez MSA, et al. (2022) Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. *Journal of Medical Virology* 94(5): 1825-1832.
21. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF (2020) The proximal origin of SARS-CoV-2. *Nat Med* 26: 450-452.
22. Almazán F, Sola I, Zuñiga S, Marquez-Jurado S, Morales L, et al. (2014) Coronavirus reverse genetic systems: Infectious clones and replicons. *Virus Research* 189: 262-270.
23. Huang X, Zhang C, Pearce R, Omenn GS, Zhang Y (2020) Identifying the Zoonotic Origin of SARS-CoV-2 by Modeling the Binding Affinity between the Spike Receptor-Binding Domain and Host ACE2. *J Proteome Res* 19: 4844-4856.
24. Lam TTY, Jia N, Zhang YW, Shum MHH, Jiang JF, et al. (2020) Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* 583: 282-285.
25. Zhang X, Chen X, Zhang Z, Roy A, Shen Y (2020) Strategies to trace back the origin of COVID-19. *Journal of Infection* 80(6): e39-e40.
26. Zhou P, Yang X-L, Wang XG, Hu B, Zhang L, et al. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579: 270-273.
27. Romeu AR, Ollé E (2021) SARS-CoV-2 Origin: An Affair of Codons? *BIOLOGY*.
28. Peacock TP, Sheppard CM, Brown JC, Goonawardane N, Zhou J, et al. (2021) The SARS-CoV-2 variants associated with infections in India, B.1.617, show enhanced spike cleavage by furin. *Microbiology, biorxiv*.

