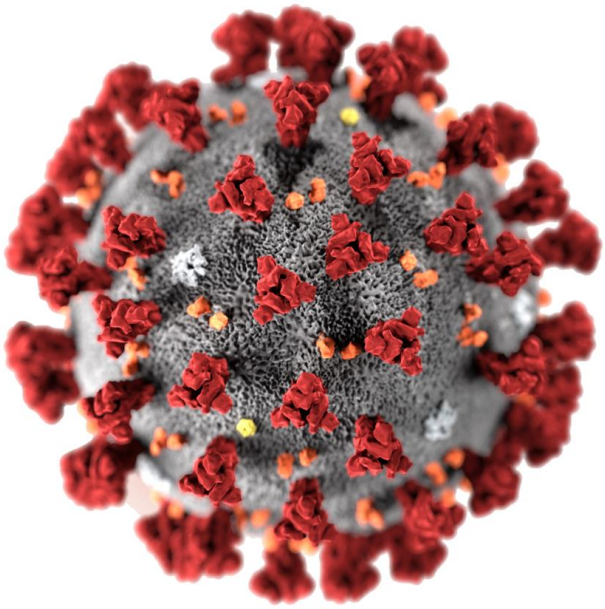
The origin of SARS-CoV-2: how, where and when did the virus emerge?



Master's Thesis by BSc. M. Schutte (6112072)

Master Infection & Immunity, Utrecht University

Supervisor: Dr. B.L. Haagmans

Examiner: Prof. dr. F.J.M. van Kuppeveld

17-12-2021

Utrecht University

LAYMEN'S SUMMARY

The ongoing pandemic of COVID-19 has been holding the whole world in its grip for almost two years. Soon after the initial reports of an unknown lung disease, the cause was identified to be a novel coronavirus called SARS-CoV-2. The pandemic has influenced billions of lives, causing millions of deaths worldwide and the numbers are still rising. In order to prevent a virus outbreak with such global impact in the future, it is crucial to gain more knowledge about how viruses emerge in the human population. If the factors that influence the introduction of novel viruses to humans are known, we might be able to stop the emergence of novel viruses by intervening with these factors. The aim of this paper is to increase the understanding of the origin of SARS-CoV-2 by providing a thorough, unbiased review of the current literature on this topic. Three aspects of the virus' origin are discussed: how, where and when this virus was introduced in the human population. Two main routes of emergence of the virus are currently being considered: through nature or in a laboratory. Previous coronavirus outbreaks in human history have been linked to animal-to-human transmission and hence have a natural origin, with bats being the most prevalent animal source. While related viruses have been found in nature, no animal source has yet been identified for SARS-CoV-2. Moreover, the unique features that are seen in the SARS-CoV-2 genome urge scientists to reconsider a laboratory origin of the virus. Several laboratories across the world have been working on coronaviruses for decades and state-of-the-art experimental techniques allow researchers to create viruses without leaving any traces of it being unnatural. Regarding the initial location and timing of the virus' emergence, the consensus is that the pandemic began in Wuhan, China in December 2019. Outside of China, however, investigations from Europe and America have reported the prevalence of the virus before that period. These reports raise the question whether SARS-CoV-2 could have originated outside of China before December 2019. Overall, it is clear that the origins of SARS-CoV-2 remain a point of discussion. The literature overview and in-depth discussion presented here forms a basis for further research into the origin of SARS-CoV-2. With this study, we hope to contribute to increasing the preparedness for future virus outbreaks to prevent our lives from ever being locked-down again.

ABSTRACT

Since the first reports of a novel pneumonia in Wuhan, China in December 2019, the coronavirus disease (COVID-19), caused by SARS-CoV-2, has become a pandemic, affecting millions of lives worldwide. Despite efforts of researchers across the globe, the origin of SARS-CoV-2 remains undetermined and there is a need for clarification on the emergence of the virus in the human population. A central aspect in this discussion is the origin of two unique genomic features of the novel virus: the receptor binding domain (RBD) of the spike protein and a polybasic insertion giving rise to a furin cleavage site. The purpose of this review is to enhance the understanding of SARS-CoV-2's emergence through an extensive literature review. In this paper, the current literature on the virus' origin is critically and objectively reviewed in order to guide future research regarding the virus' initial route of introduction, location, and timing. The information presented here can be used as a basis for further research. Future work should include sampling a wide variety of animal species across a broad geographical region. Additionally, transparency and cooperation from laboratories across the world is required. In essence, there is a need for an open and unbiased discussion amongst scientists regarding the possibilities of a natural as well as a laboratory origin of the virus. This might bring us closer to identifying the true origin of SARS-CoV-2 and create insights that will allow better preparedness for future outbreaks.

INTRODUCTION

The rapid spread of a novel virus causing respiratory illness across the globe has raised major global health concerns. Since the first cases of a novel type of pneumonia that were reported in Wuhan, China in December 2019 (1,2), the virus has caused 269 million cases and 5.3 million deaths worldwide as of December 12, 2021 (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports). The causative agent for this pandemic was identified as SARS-CoV-2, responsible for the coronavirus disease (COVID-19) (1,2). SARS-CoV-2 is a SARS-like coronavirus belonging to the Sarbecovirus subgenus of the *Betacoronaviridae* (3), which also contains SARS-CoV, the virus that caused an epidemic in 2002-2004 (4). Both viruses use angiotensin-converting enzyme 2 (ACE2), which is expressed in a variety of human tissues (5), as an entry receptor through binding with the receptor binding domain (RBD) of the transmembrane spike protein (2,6). Despite belonging to the same subgenus, SARS-CoV and SARS-CoV-2 share only 79% sequence identity (3). Phylogenetic analysis showed that the novel virus was most closely related to a coronavirus isolated from bats (*Rhinopholus affinis*) (2). This bat coronavirus, called RaTG13, showed 96.2% sequence identity with SARS-CoV-2. Although the overall sequence is similar, SARS-CoV-2 contains a few characterizing genomic features that are not shared with RaTG13. The first is the RBD, which is crucial for interaction with a host cell. The RBD of SARS-CoV-2 differs significantly from the RBD of RaTG13 (7), suggesting a large evolutionary distance between these viruses. The second unique feature of the SARS-CoV-2 genome is a polybasic insertion in the spike protein, creating a furin cleavage site (8). A furin cleavage site is also important in infection, facilitating cell entry, and may alter virus tropism (9). This insertion is not only absent in RaTG13, but also in the whole genus of betacoronaviruses, while they have been observed in more distinct human coronaviruses such as MERS-CoV and HKU1 (10,11). These unique features of SARS-CoV-2 have raised a discussion about the virus' origin, in which two main theories can be distinguished: a natural or laboratory origin (Figure 1). The broad diversity of coronaviruses in bats (12), their predisposition for recombination (13,14) and crossing species barriers (15) have led many scientists to believe that SARS-CoV-2 arose through a zoonotic jump, possibly with an intermediate host (Figure 1, 'zoonosis' and 'direct zoonosis') (16,17). Opposingly, a laboratory origin of the virus can be argued, since state-of-the-art laboratory techniques such as genetic engineering or serial passaging are theoretically able to have given rise to the unique genomic features of SARS-CoV-2 (Figure 1, 'escape or infection') (18). The virus could also have been isolated from nature, followed by escape of the wild-type virus from the laboratory (Figure 1, 'escape or infection'). In addition to the discussion on how the virus entered the human population, recent reports of early virus circulation outside of China have raised doubts about

the initial location and timing of the virus' emergence (19–28). Despite the efforts of many researchers in the field to prove the virus' origin, the debate continues to be inconclusive, and there remains a need for clarification on the origin of the virus that is causing the ongoing pandemic. This fundamental issue is worthwhile to investigate since it can aid in the containment of the current pandemic and increased preparedness for future outbreaks.

Here, the current literature on the origin of SARS-CoV-2 is critically and objectively reviewed, providing a basis for future research. The emergence of this novel coronavirus regarding its route of introduction as well as location and timing are discussed.

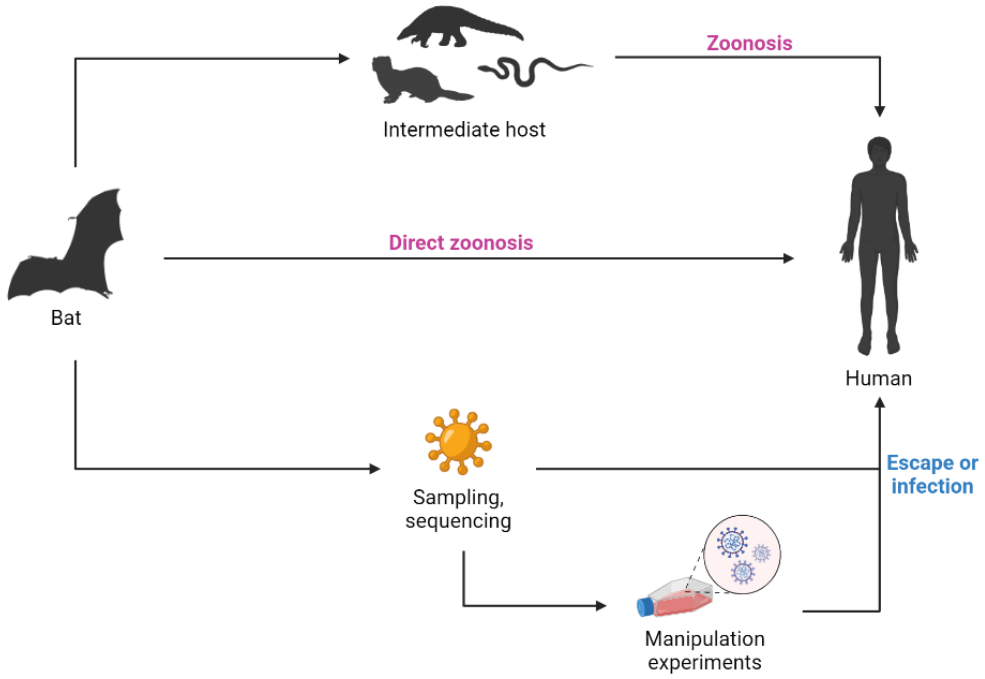


Figure 1: Potential routes of introduction of SARS-CoV-2 in the human population. A natural zoonotic jump directly from bats or through an intermediate host to humans, possibly involving recombination, is one plausible scenario (routes indicated in pink). Alternatively, virus sampling from wild bats could have resulted in human infection either directly from the sampled virus or after laboratory experiments such as genetic engineering or serial passage (indicated in blue). In the laboratory, the virus could have infected staff members, or it could have escaped the laboratory.

HOW DID SARS-COV-2 EMERGE IN THE HUMAN POPULATION?

The initial route of transmission of SARS-CoV-2 to humans fuels an ongoing discussion in which two main theories can be distinguished: a natural, zoonotic origin and a laboratory leak. Figure 1 depicts all potential routes of introduction that will be discussed in this section. Two large organizations, the World Health Organization (WHO) and the US Intelligence Community (IC), have each conducted an extensive study aiming to determine which theory on the virus' origin might be true (29,30). While the WHO published an extensive report including the methods, data and findings of their study, the IC only released a summary of their findings to the public. The WHO asserted that while related viruses have been found in bats and

pangolins, no virus identified thus far is similar enough to SARS-CoV-2 to be its direct progenitor (29). However, likewise, there is currently no substantial evidence supporting the opposing laboratory theory. In the end, the WHO assessed a natural emergence of SARS-CoV-2 through an intermediate host to be the most likely scenario; a laboratory incident was considered to be an extremely unlikely cause. The IC, on the other hand, does not seem as certain as the WHO; they described that the IC remains divided on the most likely pathway of emergence of SARS-CoV-2 (30). The IC stated that both hypotheses are plausible and new information is required to allow the determination of the virus' true origin. This statement is reiterated in the IC's updated assessment that was published a few months later (31). Overall, it is clear that the origin of the ongoing pandemic in terms of its route of emergence remains to be elucidated. As mentioned in the introduction, the central aspects in this discussion are the two unique genomic characteristics of SARS-CoV-2, being the furin cleavage site and the RBD, as well as the large evolutionary gap between its closest found relatives. In this section, both theories will be objectively discussed with regard to the aforementioned aspects.

Zoonotic jump

Given that all prior coronavirus introductions into the human population have had zoonotic origins (32), a zoonotic origin should be considered for SARS-CoV-2 as well. Of particular interest are bats, since they are a known reservoir for a wide range of viruses, including coronaviruses (33). The most recent coronavirus outbreaks in humans before the pandemic of December 2019 were caused by SARS-CoV in 2002 and MERS-CoV in 2012 and both viruses are known to have originated from bats (4,32–34). Spill-overs of bat coronaviruses to humans often occur through an intermediate host, as seen with SARS-CoV that first spilled from bats to civets (35) and MERS-CoV that first infected dromedary camels before jumping to humans (32,36). All this leads to the question whether SARS-CoV-2 could also have originated from bats and whether an intermediate host species was involved (Figure 1, 'zoonosis' and 'direct zoonosis').

Indeed, SARS-CoV-2-related coronaviruses have been found in bats. The first sequence analysis of SARS-CoV-2 showed that the virus shares as much as 96% sequence identity with the coronavirus RaTG13 from *Rhinolophus affinis* bats (2), which is the closest known relative to the novel coronavirus. While this similarity implies that bats are involved in the emergence of the novel coronavirus, the 4% genetic difference suggests a missing link. Subsequent sequence analysis studies identified two additional bat coronaviruses closely related to SARS-CoV-2. First, Zhou *et. al.* found that the bat coronavirus RmYN02 from *Rhinolophus malayanus* shares 93.3% sequence identity with SARS-CoV-2 (37). Remarkably, like SARS-CoV-2, RmYN02 contains an insertion at the S1/S2 junction site of the spike protein, providing another genomic link between SARS-CoV-2 and a bat coronavirus. The same

research group more recently identified several additional SARS-CoV-2-related viruses in bats sampled from the Yunnan province in China (14). From these viruses, RpYN06 from *Rhinolophus pusillus* was identified as the closest relative to SARS-CoV-2 after RaTG13, with 94.48% sequence identity. However, a common feature of all these SARS-CoV-2-related bat coronaviruses is their distinctively different RBD compared to SARS-CoV-2 (2,14,37). Notably, the three viruses discussed above all originate from different species of *Rhinolophus* bats. This triggered further research into coronaviruses circulating in *Rhinolophus* bat species. Two additional viruses, STT182 and STT200, from another *Rhinolophus* species known as *R. shameli* in Cambodia were recently reported to share 92.6% sequence identity with SARS-CoV-2 (38). Interestingly, these novel bat coronaviruses also share five out of six key RBD residues with SARS-CoV-2, which had not been observed before in bat coronaviruses. More recently, even more distinctive bat coronaviruses have been identified with highly similar RBDs to that of SARS-CoV-2 (39). These viruses, called BANAL-52 and BANAL-103 were isolated from *R. malayanus* and *R. pusillus*, respectively, in northern Laos, providing further evidence for circulation of SARS-CoV-2-like viruses in different bat species across Southeast Asia. All in all, these phylogenetic studies demonstrate that bats from Southeast Asia host a large variety of SARS-CoV-2-related viruses.

Despite the similarities between bat coronaviruses and SARS-CoV-2, the 4% genomic difference between the two could suggest that there might be an intermediate host that harbours a virus that is even more closely related. As mentioned before, the previous human coronavirus outbreaks of SARS-CoV and MERS-CoV also originated through bats via intermediate host species (35,36). However, since SARS-CoV-2 shares only 79% or 50% sequence identity with SARS-CoV or MERS-CoV (3), respectively, it seems unlikely that the novel coronavirus shares its intermediate host with these older viruses. An indication that SARS-CoV-2 might have an intermediate host is presented by Zhou *et. al.*, who show that SARS-CoV-2 can use human, Chinese horseshoe bat (*Rhinolophus*), civet and pig ACE2 proteins as entry receptor (2). These findings triggered more research into the potential intermediate hosts of SARS-CoV-2.

Current literature might point to Malayan pangolins as the intermediate host for SARS-CoV-2. In October 2019, just prior to the COVID-19 outbreak, Liu *et. al.* discovered SARS-like coronaviruses in samples from two dead Malayan Pangolins (*Manis javanica*) (40). This discovery led Zhang *et. al.* to dive into the possible relationship between a novel pangolin coronavirus (Pangolin-CoV) and SARS-CoV-2, and they found 91.02% sequence identity between the two (41,42). Moreover, Pangolin-CoV is 90.55% identical to RaTG13, the closest bat coronavirus relative to SARS-CoV-2. Despite the lower overall sequence identity, the RBD of Pangolin-CoV much more closely resembles that of SARS-CoV-2 than RaTG13, with five of

the six key residues being consistent. Only a single amino acid change in the RBD is observed between Pangolin-CoV and SARS-CoV-2, indicating that the two viruses share the same ACE2 receptor (43). An argument against pangolins as the intermediate host is that pangolins are infrequently in contact with humans or bats, making it unlikely for Pangolin-CoV to be the direct ancestor of SARS-CoV-2 (29). Nonetheless, pangolins seem to be a natural host of SARS-like coronaviruses (43). Altogether, these data demonstrate that Pangolin-CoV is genetically related to both SARS-CoV-2 and RaTG13, making the pangolin a potential intermediate host for SARS-CoV-2.

Other animals, in addition to pangolins, have been suggested to be potential intermediate hosts for SARS-CoV-2. Mink, ferrets, snakes, and turtles are some of the other wildlife that have been considered (16). Mink farms in the Netherlands have reported SARS-CoV-2 virus outbreaks among their animals (44). Infected minks have also been demonstrated to transmit the virus to humans. Despite the fact that ferrets are highly susceptible to SARS-CoV-2, the virus does not effectively exploit ferret ACE2 (45). Based on assumptions about the interaction between the RBD and the host ACE2, snakes and turtles, which can be found on wildlife markets, have been suggested as potential intermediate hosts (17). However, this allegation has been refuted for both animal species, since almost half of the key residues of the ACE2 from these animals has been abolished (46), making the prediction unreliable. Moreover, domestic animals such as dogs and cats should not be disregarded as candidate hosts, due to their close contacts with humans. Especially cats have shown high susceptibility to SARS-CoV-2, while dogs are only moderately susceptible (47). For a comprehensive overview of the potential intermediate hosts for SARS-CoV-2, the reader is referred to table 1 of the paper by Zhao *et. al.* (16). All in all, the true intermediate host for SARS-CoV-2 remains unidentified.

Having touched upon the potential animal sources of SARS-CoV-2, the question remains how the unique genomic features of the virus could have arisen naturally. Regarding the RBD, this feature consists of six key residues for ACE2 binding of which one is shared by RaTG13 and five are shared by Pangolin-CoV. SARS-CoV-2 binds with high affinity to ACE2 of humas but also of various species such as bats, ferrets, cats and monkeys (7). Andersen *et. al*. argued that the high affinity binding to human ACE2 is likely the result of natural selection (48). They stated that while the RBD binds human ACE2 with high affinity, the binding is not the most optimal and this unique solution could not have been predicted or manufactured. The very recent discovery of bat coronaviruses that contain a similar RBD as SARS-CoV-2 (38,39) have led to the idea that SARS-CoV-2 could have emerged through recombination of bat coronaviruses. Evidence for recombination in SARS-like bat coronaviruses has already been presented as early as 2007, following the SARS pandemic, through the identification of a

potential recombination breakpoint (49). Regarding SARS-CoV-2, Temmam *et. al.* suggest that with the sequences found in different *Rhinolophus* species including RaTG13, RmYN02, RpYN06, BANAL-52 and BANAL-103, all discussed earlier in this paper, the mosaic genome of SARS-CoV-2 can be constructed (39). This theory implies that no intermediate host would have been necessary for SARS-CoV-2 to emerge. They support their theory by identifying 14 recombinant breakpoints during the evolutionary history of sarbecoviruses. The recombinant history of sarbecoviruses has also been recognized by other researchers including Boni *et. al.*, who found various recombination breakpoints in the history of *Sarbecovirus* genomes and assessed that sarbecoviruses undergo frequent recombination (50). Furthermore, Zhu *et. al.* identified three possible recombination events that could have resulted in the origin of SARS-CoV-2 (51). In contrast to Temmam *et. al.*, who suggest that recombination occurred between bat coronaviruses (39), Zhu *et. al.* propose that the novel coronavirus possibly resulted from a recombination event between bat and pangolin coronaviruses (51). However, it should be noted that at the time of the Zhu *et. al.* study, a SARS-CoV-2-like RBD had not yet been observed in bat coronaviruses. Altogether, recombination events could be a plausible explanation for the unique genome of SARS-CoV-2.

While recombination might explain the unique RBD of SARS-CoV-2, explaining the origin of the furin cleavage site might form a greater challenge. SARS-CoV-2 contains an insertion of the 4 amino acids PRRA at the S1/S2 subunit junction of the spike protein, resulting in a polybasic cleavage site (8). Since furin, a host protease, is highly expressed in the lungs, this insertion might benefit the viral fitness by increasing the pathogenicity and infectivity (8). No other betacoronaviruses contain such an insertion, although other human coronaviruses such as MERS-CoV and HKU1 and the SARS-like bat coronavirus RmYN02 do harbour polybasic insertions (10,11,37). This provides evidence that polybasic insertions do occur naturally in coronaviruses. Therefore, a possible origin of the furin cleavage site in the SARS-CoV-2 genome is convergent evolution. Nowadays, we still see the ongoing evolution of the virus (52); from time to time, it acquires a novel mutation that increases the viral fitness (53). Moreover, since a furin cleavage site is thought to increase viral fitness (8), it is conceivable that the polybasic insertion in the SARS-CoV-2 genome resulted from convergent evolution as a way to adapt to humans as a host. Alternatively, since polybasic insertions have been observed in SARS-CoV-2-like viruses (10,11,37), recombination could have resulted in the furin cleavage site of SARS-CoV-2 as well. Essentially, while indications can be found for a natural origin of SARS-CoV-2, the currently available evidence is indirect and therefore the questions about the origin of the key genomic features of SARS-CoV-2 remain unanswered.

Laboratory leak

A laboratory origin of the novel 2019 human coronavirus cannot be abrogated. While this theory seems to have less support than the natural origin theory, it is important to consider all possible SARS-CoV-2 emergence routes. Only two research groups thoroughly evaluated the possibility of a laboratory origin of the virus (18,54). Many other studies that name a laboratory origin already dismiss the theory as improbable without objectively weighing all arguments. Here, arguments in favour and against a laboratory origin of SARS-CoV-2 will be presented.

The potential routes of introduction of the virus into the human population from a laboratory leak can be distinguished into two main scenarios (Figure 1, 'escape or infection'). One plausible research-related explanation of origin is that a virus sampled from nature that was being characterized in the laboratory accidentally escaped the laboratory. Possible scenarios for a laboratory leak are infection of a staff member through contact with an infected animal or accidental self-inoculation, failure in equipment or escape through the sewage system (18). Alternatively, the unique genomic features of SARS-CoV-2 lead to the speculation that the virus might have originated in the laboratory through state-of-the-art experimental techniques such as genetic engineering or gain-of-function studies (18,54). The plausibility of a laboratory origin of SARS-CoV-2 is exemplified by the fact that laboratory leaks have occurred in the past. For instance, in 1977, research on the H1N1 influenza virus, which was already decades old at the time, caused a pandemic through a laboratory leak of the virus (55). Thus, regardless of the exact route of escape, a laboratory origin of SARS-CoV-2 should seriously be considered.

Furthermore, Segreto and Deigin assert that the insertion in SARS-CoV-2 cannot be natural since it is out of frame when comparing the sequence to RaTG13 and Pangolin-CoV (18). Therefore, the insertion could not have been caused by polymerase slippage or by releasing and repriming, since these mechanisms are thought to maintain the reading frame. Opposingly, a natural explanation for the out-of-frame insertion is provided by Holmes *et. al.* (53). They state that furin cleavage sites are common in other coronavirus spikes; a nearly identical sequence is even found in the bat coronavirus HKU9-1. Both HKU9-1 and SARS-CoV-2 contain a short upstream sequence that indicates natural recombination. Therefore, an out-of-frame insertion of a furin cleavage site can be explained by evolutionary mechanisms. Still, this notion does not form direct evidence for a natural origin, leaving the possibility of a laboratory leak open.

Since the COVID-19 outbreak of December 2019 started in Wuhan, the Wuhan Institute of Virology (WIV) is being considered as a potential source of the virus (18). The WIV has been performing coronavirus research for almost two decades. However, in 2018, safety concerns

about the WIV's laboratory were raised (18). It was stated that laboratory materials were not properly discarded, so that infectious material could end up in the regular trash can or the sewage system. Moreover, the WIV has worked on the RaTG13 virus, the closest known relative of SARS-CoV-2 (29). Notably, the laboratory had already sequenced the RaTG13 genome in 2018, but only published it after the identification of SARS-CoV-2 (18,56). Another discriminating observation is the move of the Wuhan CDC laboratory on 2 December 2019 to a location near the Huanan market (29). While this may seem like an unfortunate coincidence, a link to the virus outbreak cannot be ruled out. Nonetheless, no incidents associated to the move were reported and no staff of the WIV was reported to be infected in December 2019 (29). Importantly, besides the WIV, other laboratories outside of China have long been working on coronaviruses as well and a laboratory escape could just as well have originated from one of those laboratories. Overall, there is not enough evidence pointing at the WIV as the definitive source of a laboratory escape as of now, leaving all possibilities of a laboratory escape open.

The coronavirus research, including gain-of-function experiments and genetic engineering, that has been going on for the past 20 years across the world, could in theory explain the unique genomic features that have been observed in the SARS-CoV-2 genome. As mentioned before, SARS-CoV-2 contains a unique RBD and furin cleavage site. While these features have both been observed in nature in related viruses (37–39,41), laboratory techniques might also be able to explain the emergence of these unique features. One of the experimental techniques that is often used to study the pandemic potential of viruses is the production of chimeric viruses, often using bat coronaviruses as backbones, augmented with a spike protein able to bind human ACE2 (54). By combining naturally occurring coronaviruses that display either the unique RBD or the furin cleavage site, theoretically, this could have resulted in the virus that we know as SARS-CoV-2. Interestingly, bat-human chimeric coronaviruses have been produced by the Shi group from the WIV in 2007 and again in 2015 and 2017 (57,58). Another research group from the University of North Carolina also experimented with chimeric coronaviruses in 2008 and collaborated with the Shi group in 2015 (58,59). This demonstrates the existence of SARS-like chimeric viruses in different laboratories with the ability to bind human ACE2. However, it should be noted that no virus closely related to SARS-CoV-2 was studied in any laboratory before December 2019 (29). Moreover, genetic data on SARS-CoV-2 show that the virus was not derived from a previously used virus backbone (48). Nonetheless, it is evident that viruses with pandemic potential can be manufactured in the laboratory, risking a laboratory escape.

Another experimental technique that potentially created the unique RBD and furin cleavage site of SARS-COV-2 is serial passage. Serial passage is the passaging of a virus through cells or animals, forcing the same molecular adaptations that occur in a natural zoonotic jump (54).

Put simply, serial passage causes accelerated evolution of a virus. Since SARS-CoV-2 is significantly distant from other viruses, it can be argued that this novel virus could have originated from the accelerated evolution forced by serial passage. During serial passage, mutations in the RBD could have been selected for, resulting in the unique RBD of SARS-CoV-2 (54). A counterargument is that the virus, since the emergence of the initial isolates, has acquired new mutations once in a while, indicating further adaptation to humans (53). This observation refutes the claim that the spike protein was already optimized for binding human ACE2 from the beginning. Nonetheless, without further evidence, it cannot be excluded that a laboratory technique might have given rise to the unique RBD of SARS-CoV-2. Regarding the furin cleavage site, it is important to acknowledge that serial passage experiments in the past have resulted in the acquisition of a furin cleavage site by viruses (54,60). Through serial passaging, avian influenza virus and Newcastle Disease Virus have acquired a (similar) cleavage site. The polybasic insertion increased the pathogenicity of avian influenza viruses by expanding the tissue tropism (60). Since this insertion is not seen in other betacoronaviruses and other viruses have been shown to acquire it in the laboratory, it is plausible that SARS-CoV-2 also acquired its unique furin cleavage site in the laboratory.

Serial passage can be conducted in cells or animals. Considering the high susceptibility of ferrets and mink to SARS-CoV-2 (44,45), these animals could possibly have been used for serial passage to create the novel coronavirus. This is supported by the observation that SARS-CoV-2 can be transmitted from humans to mink and vice versa (44). However, an opposing argument is that if SARS-CoV-2 was passaged through such an animal model, it would likely have acquired mutations for adaptation to that species; yet, no such mutations have been found in early isolates of the pandemic virus (53). Therefore, a more probable alternative is that serial passage has been conducted in humanized mice (54,61,62). In this humanized animal model, the virus could have been passaged to force accelerated evolution, while allowing the virus to adapt to humans. Accordingly, the possibility that SARS-CoV-2 originated using experimental techniques such as serial passage remains probable.

WHERE AND WHEN WAS SARS-COV-2 INTRODUCED INTO THE HUMAN POPULATION?

The question of how SARS-CoV-2 emerged in the human population is accompanied by the question of where and when the virus arose. The first cases of a novel type of pneumonia, later identified as COVID-19, were reported in Wuhan, China in December 2019 (2,63). Since then, reports of COVID-19 quickly emerged across the globe, marking the onset of the pandemic. However, recently, indications for pre-pandemic circulation of SARS-CoV-2 outside

of China have arisen. In this section, the potential locations and timing of the novel coronavirus' introduction in the human population will be discussed.

Several approaches can be employed to detect the presence of a virus at a certain location or time. One involves nucleic acid-based methods, of which the polymerase chain reaction (PCR) is the most widely used (64). More specifically, RT-qPCR is often employed since it measures the number of amplicons, allowing quantification of the viral RNA in the sample. Additionally, this method is generally highly specific and sensitive, using a specific probe that needs to bind the target sequence. Alternatively, immunoassay-based methods can be applied (64). Immunoassay-based methods make use of antibodies to detect viral epitopes in a sample. The most widely used immunoassay is an enzyme-linked immunosorbent assay (ELISA). This method can be used to detect viral antigens in a sample using available antibodies or to detect the seroprevalence against a certain virus. In the latter case, patients can be identified who have previously been exposed to the virus based on the presence of virus-specific host-produced antibodies in their serum. A challenge associated with this type of assay is cross-reactivity. If a patient has previously been exposed to a related virus, those pre-existing instead of novel antibodies might cross-react with the antigen that you are screening with, resulting in a false positive sample. In addition to nucleic acid- and immunoassay-based techniques, the more recently developed method of next-generation sequencing (NGS) can be employed. Using NGS, potential animal hosts of viruses can be routinely sampled and tested to determine the virus prevalence. This surveillance of viral prevalence in potential animal reservoirs could aid in the prediction and prevention of a viral outbreak. To summarize, the main methods to detect a virus in a sample are nucleic acid-based methods such as PCR and immunoassay-based methods such as ELISA, possibly complemented with NGS.

In search of the location where SARS-CoV-2 first emerged, a link was found between the hospitalized patients with confirmed SARS-CoV-2 infection in Wuhan and exposure to the Huanan seafood market (63). This market, located in Wuhan, sells animals and animal products to the public and it was suggested that this could have been the origin of a zoonotic spread of SARS-CoV-2. Aiming to gather more information on the virus origin, the WHO conducted a study in China from 14 January to 10 February 2021 with an international team of experts (29). In this study, samples from the Huanan market were tested for presence of the virus using PCR and sequencing. While the team found no evidence for infections of animals on the Huanan market, the virus was found on packages of cold-chain products that were supplied to China by other countries. Further epidemiological analysis showed that the first COVID-19 cases all had been exposed to cold chain products. These findings suggest the import of cold chain products as a possible route of early virus spreading.

Although Huang *et. al.* in their study found that as much as 66% of the patients had been exposed to the Huanan market (63), a large proportion of the first cases have no ties to this market (65), leaving the possibility that the Huanan seafood market was not the original location of the SARS-CoV-2 outbreak, but perhaps merely a superspreading event (29). Sequencing data demonstrated that in the early phases of the outbreak in Wuhan, the virus isolates already showed diversity, indicating unknown transmission chains beyond the Huanan seafood market (29). Focussing more broadly on the Hubei province, this region hosts a great diversity of bat species that all carry an extensive variety of coronaviruses (12,33,50). Keeping in mind the previous coronavirus outbreaks in humans of SARS and MERS that originated from bats through an intermediate host (33), it is probable that the current outbreak also finds its origin in bats. With the widespread occurrence of bats and their diverse set of coronaviruses in the region surrounding Wuhan (12), a zoonotic spill-over from bats could have occurred close to or in Wuhan. Moreover, in Wuhan, research institutes such as the WIV have been studying coronaviruses for decades already (2), providing another link between coronaviruses and the city of Wuhan.

Despite reports of the initial COVID-19 cases and the several additional links between coronaviruses and Wuhan, there are now indications that the virus was already circulating outside of China before December 2019. Using the immunoassay-based method ELISA, studies from several countries reported the discovery of specific anti-SARS-CoV-2 antibodies in serum samples dating back to the end of 2019 (20,24,26,28). A study in Italy found seropositive samples from as early as September 2019, suggesting a pre-pandemic circulation of the virus in Italy (20). A large serum sample study in France found anti-SARS-CoV-2 antibodies from between November 2019 and January 2020 (24). In addition to these European countries, serological testing also identified positive pre-pandemic samples in the USA and Brazil (26,28). Furthermore, the nucleic acid-based technique RT-PCR has been applied by several countries on sewage samples from before or during the early stages of the pandemic (21–23,27). Sewage analysis forms a valuable tool to monitor virus circulation (66), since asymptomatic people already shed the virus through faeces, allowing a timely detection. Indeed, Peccia *et. al.* demonstrate that the SARS-CoV-2 virus concentrations in the sewage system follow the COVID-19 epidemiological curve (66). It should nonetheless be noted that currently no standard protocol for sewage surveillance is in use. Still, the detection of viral RNA in the sewage system in early December 2019 might provide strong evidence for circulation of the virus outside of China before the outbreak. Studies from Italy, Spain, the Netherlands and Brazil that used RT-PCR to analyse their pre- or early pandemic sewage samples report positive samples from before the first COVID-19 cases were identified in these countries, suggesting an early circulation of the virus outside of China (21–23,27). Additionally, the

method of RT-PCR was applied to respiratory and oropharyngeal samples from France and Italy, respectively, providing additional evidence for an early virus circulation in these countries (19,25). While most of the studies discussed here mention limitations to their work such as sample quality and technique sensitivity, these early indications for SARS-CoV-2 outside China are alarming and raise the awareness to consider countries outside China as the possible home country of the virus.

DISCUSSION

Despite the efforts of researchers worldwide, the origin of SARS-CoV-2 is still unidentified. Researchers have been unsuccessful in proving either a natural or laboratory origin of SARS-CoV-2, and so the discussion remains undecided. Moreover, research into the initial location and timing of the virus' emergence has raised additional questions. In this review, we present an objective overview of the current literature on the origin of SARS-CoV-2 and discuss future outlooks.

Regarding the route of emergence of SARS-CoV-2, two main theories are being considered: a natural, zoonotic origin and a laboratory origin. In spite of the efforts of many researchers worldwide to determine the true origin of SARS-CoV-2, now, almost two years after the onset of the pandemic, neither a zoonosis nor a laboratory leak has been proven to be true. While related viruses have been found in nature (37–39,41), the evolutionary distance is still rather large, leaving the natural ancestor of the virus yet to be identified. However, the unique genomic features of the novel virus have been found in those related viruses, leading to the possibility that SARS-CoV-2 originated through a natural process such as recombination (39,49–51). Alternatively, these genomic features could just as well have arisen in a laboratory, for example through serial passage or genetic engineering (18,54). Thus, while arguments can be given in favour of both theories, both theories are missing direct evidence, according to van Helden *et. al.* (62). The discovery of related viruses in nature does not necessarily mean that the virus has a natural origin; these related viruses could just as well have been used in the laboratory to design the novel virus. Eventually, no direct ancestor or natural host has yet been identified. Furthermore, despite the list of potential intermediate hosts (16), the true intermediate host of SARS-CoV-2 has not been discovered. Moreover, although present-day laboratory techniques allow researchers to create novel viruses without leaving a trace of being non-natural, the fact that this is possible does not signify a laboratory origin to be true. With the evidence currently available, neither theory can be proven to be correct.

The theory of a zoonotic origin of the virus, while being the most widely accepted, lacks direct evidence. Previous research has sampled an insufficient number of animal species and across too narrow regions. Research has mainly focussed on bats, while the true natural host

of SARS-CoV-2 might not have been sampled yet due to this restricted focus of researchers. Further research is required to shed more light on the route of emergence of SARS-CoV-2 in the human population. As advised by the WHO and suggested by researchers in the field, to either prove or abrogate the theory of a natural origin, global investigations are needed to sample a large variety of potential animal hosts across a broad geographical region to detect the presence of SARS-CoV-2-like viruses (29,67). Furthermore, in order to prevent future zoonotic spill-overs, we need to limit human contributions to spill-over events. Burki reports that human behaviour such as deforestation, intensive farming and bat cave tourism are contributors to zoonoses and need to be limited (68). Prof. James Wood of the University of Cambridge adds that biodiversity loss is another driver of zoonoses and that we should therefore focus on biodiversity protection (69). He further argues that wild animal markets should be more tightly regulated or even banned.

The hypothesis of the virus's origin in a laboratory has generally received little consideration. This notion of the virus's origin was widely regarded as a conspiracy theory in the first year after the pandemic (70), resulting in scientists not seriously considering it. However, almost another year after the emergence of the novel coronavirus, the opposite theory of a natural origin has not been proven to be true, provoking scientists to revisit the possibility of the laboratory origin. The American IC concluded that no definitive explanation for the origin of the virus can be ruled unless new information arises (31). Moreover, The WHO, after condemning the laboratory theory as ‘extremely unlikely’ (29), mitigated their statement after realising that their study in China insufficiently considered a laboratory scenario. Several renowned scientists now call for an open, unbiased investigation into a potential laboratory origin of SARS-CoV-2 (62,70). Such an investigation would require international collaboration to explore all possible laboratory sources of a virus escape. The assessment of the IC stresses that cooperation from China is required in order to increase our understanding of the virus’ origin (31). Up to this point, China has not been that cooperative. For instance, when the WHO team went to visit China to study the virus’ origin, their team of international experts was only allowed a few hours of supervised access to the WIV (29). Consequently, the possibility of the virus’ escape from the WIV could not be fully explored. Additional incriminating observations about the WIV are their move just prior to the outbreak, safety concerns, unshared laboratory notebooks and even deleted sequences from bat coronaviruses (18,56). The Chinese need to be transparent and comply with an open investigation into their laboratories in order to draw the suspicion away from them, according to the American National Institute of Health (NIH) (70). Moreover, the focus should not be on Chinese laboratories alone, but on laboratories in other countries as well since the initial location of the virus’ emergence is also still undecided. Looking beyond the current pandemic, van Helden *et. al.* plead that it is critical to continue the

debate about the risks of gain-of-function experiments in order to prevent future outbreaks (62).

The other topic of debate regarding the origin of SARS-CoV-2 is where and when the virus first emerged. Using diverse methods including serum screening and sewage analysis, several research groups from across the world have found indications for early circulation of SARS-CoV-2 outside of China (19–28). Pre- or early pandemic SARS-CoV-2-positive samples have been discovered in Italy, France, Spain, the Netherlands, the USA, and Brazil. Along with the observation of the WHO that the virus can persist on the packaging of cold-chain products (29), these findings lead to the question whether the virus could have originated elsewhere than Wuhan, the site where the first COVID-19 cases were reported. History teaches us that this phenomenon, an epidemic occurring at a different location than the original virus introduction, is indeed plausible, since this was the case for HIV epidemics in the past (71). Similarly, van Dorp *et. al.* found based on evolutionary sequence analysis that the virus must have jumped to humans between 6 October 2019 and 11 December 2019 and thus circulated among humans prior to the outbreak in Wuhan (72). Therefore, a reasonable scenario for the emergence of SARS-CoV-2 is that the virus circulated on cold-chain products for a while, being transported from and to Wuhan across the world, resulting in a pre-pandemic period of cryptic virus circulation. In this scenario, the pre-pandemic virus must have been less adapted to humans and less pathogenic, causing mild or asymptomatic infections that went undetected. After the acquisition of further adaptation mutations, the virus might have been able to spread more rapidly and cause more severe infections, eventually causing the outbreak in Wuhan. At the moment, we cannot contend to what extent this theory might be true.

Notably, the studies that reported pre-pandemic presence of SARS-CoV-2 outside of China (19–28) have limitations that should be acknowledged. First, most of these studies handled old, frozen samples that had to be thawed before analysis. However, the thawing process could have affected the sample quality; RNA, for instance, is quite fragile and prone to degradation upon thawing (19). Moreover, while some studies analysed similar types of samples, they all employed slightly different methods. For the analysis of viral RNA in the sewage system, for instance, no standard method is available, leading to incomparable results of the studies that analysed the presence of SARS-CoV-2 in the sewage. It is important to note that the absence of standard methods have led to discrepancies in the cut-off values that were used to determine the positive samples. For instance, Carrat *et. al.* admit that they used manufacturer-defined cut-off values for their ELISA, although a different cut-off value would have been 100% specific, only this value would have selected less samples as positive (24). Hence, these discrepancies between the studies make them incomparable, exemplifying the importance of using correct and consistent cut-off values. Furthermore, the employed methods

might lack sensitivity or specificity due to the absence of one ideal method. Taking serological analysis as an example, some samples might be false positive due to cross-reactivity; the antibodies that are detected to be reactive against SARS-CoV-2 might actually be antibodies against a common human coronavirus. Insensitive methods might not be able to discriminate between cross-reactive antibodies and truly positive samples. Altogether, limitations like these depreciate the reliability of the studies. Nonetheless, the presented evidence is intriguing and requires further investigation, as advised by the WHO (29,67).

In order to determine the true location and timing of the origin of SARS-CoV-2, there is a need for validated, standardized methods. Using standardized methods including guidelines for sample quality and standard cut-off values, laboratories across the world need to retrospectively analyse their country's pre-pandemic samples, such as serum and sewage samples. International cooperation is required to compare the results gathered on early virus circulation by different countries. To make an objective and unbiased assessment about the original location and timing of the virus, independent laboratories with no conflict of interest could be employed to conduct the analyses of pre-pandemic samples from across the world. Furthermore, the development of standardized methods for sewage and serum sample analysis could advance into a standard monitoring procedure that allows early detection of future virus outbreaks.

Few studies have approached the question of SARS-CoV-2's origin objectively and unbiased; reviews on the subject frequently focus on either the natural or laboratory origin theory. The novelty of this review lies in its objective approach to describing the current knowledge on the virus' origin, serving as a foundation for future research. With the currently available evidence, neither the theory of a natural nor that of a laboratory origin of SARS-CoV-2 can be proven or abrogated. To move forward in the search for the origin of SARS-CoV-2, there is a pressing need for open discussions and investigations (62), leaving room for either of the opposing theories on how the virus emerged to possibly be true. Such an open investigation would include all previously mentioned suggestions, including sampling a wide variety of animal species across a broad geographical region and transparency of laboratories. Nevertheless, it is vital to remember that proving either theory to be correct requires immaculate proof, which means we may never be able to confirm the virus' origin. The knowledge gathered thus far has already led to new ideas about how to reduce the risks of future outbreaks. Further research including international cooperation may extend this knowledge in order to increase the worldwide preparedness for future pandemics.

REFERENCES

1. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. Nat 2020 5797798. 2020 Feb 3. Available from: https://www.nature.com/articles/s41586-020-2008-3

2. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nat 2020 5797798. 2020 Feb 3;579(7798):270–3. Available from: https://www.nature.com/articles/s41586-020-2012-7

3. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020 Feb 22;395(10224):565–74. Available from: http://www.thelancet.com/article/S0140673620302518/fulltext

4. Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S, et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. N Engl J Med. 2003;348(20):1967–76.

5. Hikmet F, Méar L, Edvinsson Å, Micke P, Uhlén M, Lindskog C. The protein expression profile of ACE2 in human tissues. Mol Syst Biol. 2020;16(7):1–16.

6. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020 Apr 16;181(2):281-292.e6. Available from: http://www.cell.com/article/S0092867420302622/fulltext

7. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J Virol. 2020 Mar 17;94(7).

8. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res. 2020;176(February):104742. Available from: https://doi.org/10.1016/j.antiviral.2020.104742

9. Stout AE, Millet JK, Stanhope MJ, Whittaker GR. Furin cleavage sites in the spike proteins of bat and rodent coronaviruses: Implications for virus evolution and zoonotic transfer from rodent species. One Heal. 2021 Dec 1;13:100282.

10. Millet JK, Whittaker GR. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proc Natl Acad Sci. 2014 Oct 21;111(42):15214–9. Available from: https://www.pnas.org/content/111/42/15214

11. Chan C-M, Woo PCY, Lau SKP, Tse H, Chen H-L, Li F, et al. Spike Protein, S, of Human Coronavirus HKU1: Role in Viral Life Cycle and Application in Antibody Detection: https://doi.org/103181/0806-RM-197. 2017 Aug 21;233(12):1527–36. Available from: https://journals.sagepub.com/doi/10.3181/0806-RM-197

12. XD L, W W, ZY H, ZX W, WP G, XQ G, et al. Extensive diversity of coronaviruses in bats from China. Virology. 2017 Jul 1;507:1–10. Available from: https://pubmed.ncbi.nlm.nih.gov/28384506/

13. Zhang YZ, Holmes EC. A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. Cell. 2020 Apr 16;181(2):223–7.

14. Zhou H, Ji J, Chen X, Bi Y, Li J, Wang Q, et al. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. Cell. 2021 Aug 19;184(17):4380-4391.e14.

15. Chan JFW, To KKW, Tse H, Jin DY, Yuen KY. Interspecies transmission and emergence of novel viruses: lessons from bats and birds. Trends Microbiol. 2013 Oct 1;21(10):544–55.

16. Zhao J, Cui W, Tian B. The Potential Intermediate Hosts for SARS-CoV-2. Front Microbiol. 2020 Sep 30;0:2400.

17. Liu Z, Xiao X, Wei X, Li J, Yang J, Tan H, 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 Jun 1;92(6):595–601. Available from: https://onlinelibrary-wiley-com.proxy.library.uu.nl/doi/full/10.1002/jmv.25726

18. Segreto R, Deigin Y. The genetic structure of SARS-CoV-2 does not rule out a laboratory origin. BioEssays. 2021 Mar 1;43(3):2000240. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/bies.202000240

19. Amendola A, Bianchi S, Gori M, Colzani D, Canuti M, Borghi E, et al. Evidence of SARS-CoV-2 RNA in an Oropharyngeal Swab Specimen, Milan, Italy, Early December 2019. Emerg Infect Dis. 2021 Feb 1;27(2):648. Available from: /pmc/articles/PMC7853584/

20. Apolone G, Montomoli E, Manenti A, Boeri M, Sabia F, Hyseni I, et al. Unexpected detection of SARS-CoV-2 antibodies in the prepandemic period in Italy; 2020 Nov 11; Available from: https://journals.sagepub.com/doi/full/10.1177/0300891620974755

21. La Rosa G, Mancini P, Bonanno Ferraro G, Veneri C, Iaconelli M, Bonadonna L, et al. SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. Sci Total Environ. 2021;750(December 2019):141711. Available from: https://doi.org/10.1016/j.scitotenv.2020.141711

22. Chavarria-Miró G, Anfruns-Estrada E, Guix S, Paraira M, Galofré B, Sánchez G, et al. Sentinel surveillance of SARS-CoV-2 in wastewater anticipates the occurrence of COVID-19 cases. medRxiv. 2020 Jun 13;2020.06.13.20129627. Available from: https://www.medrxiv.org/content/10.1101/2020.06.13.20129627v1

23. Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. Environ Sci Technol Lett. 2020 Jul 14;7(7):511–6. Available from: https://pubs.acs.org/doi/abs/10.1021/acs.estlett.0c00357

24. Carrat F, Figoni J, Henny J, Desenclos J-C, Kab S, de Lamballerie X, et al. Evidence of early circulation of SARS-CoV-2 in France: findings from the population-based “CONSTANCES” cohort. Eur J Epidemiol 2021 362. 2021 Feb 6;36(2):219–22. Available from: https://link.springer.com/article/10.1007/s10654-020-00716-2

25. Deslandes A, Berti V, Tandjaoui-Lambotte Y, Alloui C, Carbonnelle E, Zahar JR, et al. SARS-CoV-2 was already spreading in France in late December 2019. Int J Antimicrob Agents. 2020;55(6):106006. Available from: https://doi.org/10.1016/j.ijantimicag.2020.106006

26. Basavaraju S V, Patton ME, Grimm K, Rasheed MAU, Lester S, Mills L, et al. Serologic Testing of US Blood Donations to Identify Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)–Reactive Antibodies: December 2019–January 2020. Clin Infect Dis. 2021 Jun 15;72(12):e1004–9. Available from: https://academic.oup.com/cid/article/72/12/e1004/6012472

27. Fongaro G, Stoco PH, Souza DSM, Grisard EC, Magri ME, Rogovski P, et al. SARS-CoV-2 in human sewage in Santa Catalina, Brazil, November 2019. medRxiv. 2020 Jun 29;2020.06.26.20140731. Available from: https://www.medrxiv.org/content/10.1101/2020.06.26.20140731v1

28. Stringari LL, Souza MN de, Junior NF de M, Goulart JP, Giuberti C, Dietze R, et al. Covert cases of Severe Acute Respiratory Syndrome Coronavirus 2: An obscure but present danger in regions endemic for Dengue and Chikungunya viruses. PLoS One. 2021 Jan 1;16(1):e0244937. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0244937

29. World Health Organization. WHO-convened Global Study of Origins of SARS-CoV-2 : China Part (14 January-10 February 2021). Jt WHO-China Study Team Rep. 2021;(February):120.

30. National Intellicence Council. US intelligence report on the origins of COVID-19. 2019;52(1):1–5.

31. National Intellicence Council. Updated Assessment on COVID-19 Origins.

32. Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 2018 173. 2018 Dec 10;17(3):181–92. Available from: https://www.nature.com/articles/s41579-018-

0118-9

33. Hu B, Ge X, Wang L-F, Shi Z. Bat origin of human coronaviruses. Virol J 2015 121. 2015 Dec 22;12(1):1–10. Available from: https://virologyj.biomedcentral.com/articles/10.1186/s12985-015-0422-1

34. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. http://dx.doi.org/101056/NEJMoa1211721. 2012 Nov 7;367(19):1814–20. Available from: https://www.nejm.org/doi/full/10.1056/NEJMoa1211721

35. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. Science (80-). 2003;302(5643):276–8.

36. Azhar EI, El-Kafrawy SA, Farraj SA, Hassan AM, Al-Saeed MS, Hashem AM, et al. Evidence for Camel-to-Human Transmission of MERS Coronavirus. http://dx.doi.org/101056/NEJMoa1401505. 2014 Jun 25;370(26):2499–505. Available from: https://www.nejm.org/doi/full/10.1056/nejmoa1401505

37. Zhou H, Chen X, Hu T, Li J, Song H, Liu Y, et al. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. Curr Biol. 2020 Jun 8;30(11):2196-2203.e3.

38. Hul V, Delaune D, Karlsson EA, Hassanin A, Tey PO, Baidaliuk A, et al. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. bioRxiv. 2021 Jan 26;2021.01.26.428212. Available from: https://www.biorxiv.org/content/10.1101/2021.01.26.428212v1

39. Temmam S, Vongphayloth K, Salazar EB, Munier S, Bonomi M, Régnault B, et al. Coronaviruses with a SARS-CoV-2-like receptor- binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula. Res Sq. 2021;(September 17th). Available from: https://doi.org/10.21203/rs.3.rs-871965/v1

40. Liu P, Chen W, Chen J-P. Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (Manis javanica). Viruses 2019, Vol 11, Page 979. 2019 Oct 24;11(11):979. Available from: https://www.mdpi.com/1999-4915/11/11/979/htm

41. Zhang T, Wu Q, Zhang Z. Pangolin homology associated with 2019-nCoV. bioRxiv. 2020 Feb 20;2020.02.19.950253. Available from: https://www.biorxiv.org/content/10.1101/2020.02.19.950253v1

42. Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr Biol. 2020 Apr 6;30(7):1346-1351.e2.

43. Liu P, Jiang J-Z, Wan X-F, Hua Y, Li L, Zhou J, et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLOS Pathog. 2020 May 1;16(5):e1008421. Available from: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008421

44. Munnink BBO, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science (80-). 2021 Jan 8;371(6525):172–7. Available from: https://www.science.org/doi/abs/10.1126/science.abe5901

45. Conceicao C, Thakur N, Human S, Kelly JT, Logan L, Bialy D, et al. The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. PLOS Biol. 2020 Dec 21;18(12):e3001016. Available from: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001016

46. Luan J, Jin X, Lu Y, Zhang L. SARS-CoV-2 spike protein favors ACE2 from Bovidae and Cricetidae. J Med Virol. 2020 Sep 1;92(9):1649–56. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.25817

47. Shi J, Wen Z, Zhong G, Yang H, Wang C, Huang B, et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. Science (80-). 2020 May 29;368(6494):1016–20. Available from:

https://www.science.org/doi/abs/10.1126/science.abb7015

48. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020 264. 2020 Mar 17;26(4):450–2. Available from: https://www.nature.com/articles/s41591-020-0820-9

49. Hon C-C, Lam T-Y, Shi Z-L, Drummond AJ, Yip C-W, Zeng F, et al. Evidence of the Recombinant Origin of a Bat Severe Acute Respiratory Syndrome (SARS)-Like Coronavirus and Its Implications on the Direct Ancestor of SARS Coronavirus. J Virol. 2008 Feb 15;82(4):1819. Available from: /pmc/articles/PMC2258724/

50. Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry BW, Castoe TA, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol 2020 511. 2020 Jul 28;5(11):1408–17. Available from: https://www.nature.com/articles/s41564-020-0771-4

51. Zhu Z, Meng K, Meng G. Genomic recombination events may reveal the evolution of coronavirus and the origin of SARS-CoV-2. Sci Reports 2020 101. 2020 Dec 10;10(1):1–10. Available from: https://www.nature.com/articles/s41598-020-78703-6

52. Singh D, Yi S V. On the origin and evolution of SARS-CoV-2. Exp Mol Med 2021 534. 2021 Apr 16;53(4):537–47. Available from: https://www.nature.com/articles/s12276-021-00604-z

53. Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, Wertheim JO, et al. The origins of SARS-CoV-2: A critical review. Cell. 2021 Sep 16;184(19):4848–56. Available from: http://www.cell.com/article/S0092867421009910/fulltext

54. Sirotkin K, Sirotkin D. Might SARS-CoV-2 Have Arisen via Serial Passage through an Animal Host or Cell Culture? BioEssays. 2020 Oct 1;42(10):2000091. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/bies.202000091

55. Rozo M, Gronvall GK. The reemergent 1977 H1N1 strain and the gain-of-function debate. MBio. 2015 Aug 11;6(4).

56. Bostickson B. DBs Taken Offline. 2021;2012(February):0–25.

57. Hu B, Zeng L-P, Yang X-L, Ge X-Y, Zhang W, Li B, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLOS Pathog. 2017 Nov 1;13(11):e1006698. Available from: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1006698

58. Menachery VD, Yount BL, Debbink K, Agnihothram S, Gralinski LE, Plante JA, et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nat Med 2015 2112. 2015 Nov 9;21(12):1508–13. Available from: https://www.nature.com/articles/nm.3985

59. Becker MM, Graham RL, Donaldson EF, Rockx B, Sims AC, Sheahan T, et al. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. Proc Natl Acad Sci. 2008 Dec 16;105(50):19944–9. Available from: https://www.pnas.org/content/105/50/19944

60. Tscherne DM, García-Sastre A. Virulence determinants of pandemic influenza viruses. J Clin Invest. 2011 Jan 4;121(1):6–13. Available from: http://www.jci.org

61. Segreto R, Deigin Y, McCairn K, Sousa A, Sirotkin D, Sirotkin K, et al. Should we discount the laboratory origin of COVID-19? Environ Chem Lett 2021 194. 2021 Mar 25;19(4):2743–57. Available from: https://link.springer.com/article/10.1007/s10311-021-01211-0

62. Helden J van, Butler CD, Achaz G, Canard B, Casane D, Claverie J-M, et al. An appeal for an objective, open, and transparent scientific debate about the origin of SARS-CoV-2. Lancet. 2021 Oct 16;398(10309):1402–4. Available from: http://www.thelancet.com/article/S0140673621020195/fulltext

63. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497–506.

64. Cassedy A, Parle-McDermott A, O'Kennedy R. Virus Detection: A Review of the Current and Emerging Molecular and Immunological Methods. Front Mol Biosci. 2021;8(April):1–21.

65. Chowdhury SD, Oommen AM. Epidemiology of COVID-19. J Dig Endosc. 2020 Mar;11(1):3. Available from: /pmc/articles/PMC7364648/

66. Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, et al. SARS-CoV-2 RNA concentrations in primary municipal sewage sludge as a leading indicator of COVID-19 outbreak dynamics. medRxiv. 2020 Jun 12;2020.05.19.20105999. Available from: https://www.medrxiv.org/content/10.1101/2020.05.19.20105999v2

67. Wu Z, Jin Q, Wu G, Lu J, Li M, Guo D, et al. SARS-CoV-2's origin should be investigated worldwide for pandemic prevention. Lancet. 2021 Oct 9;398(10308):1299–303. Available from: http://www.thelancet.com/article/S0140673621020201/fulltext

68. Burki T. The origin of SARS-CoV-2. Lancet Infect Dis. 2020 Sep 1;20(9):1018–9. Available from: http://www.thelancet.com/article/S1473309920306411/fulltext

69. expert reaction to a preprint reviewing the evidence on the origins of SARS-CoV-2 | Science Media Centre. Available from: https://www.sciencemediacentre.org/expert-reaction-to-a-preprint-reviewing-the-evidence-on-the-origins-of-sars-cov-2/

70. Thacker PD. The covid-19 laboratory leak hypothesis: did the media fall victim to a misinformation campaign? BMJ. 2021 Jul 8;374. Available from: https://www.bmj.com/content/374/bmj.n1656

71. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. Cold Spring Harb Perspect Med. 2011 Sep 1;1(1):a006841. Available from: http://perspectivesinmedicine.cshlp.org/content/1/1/a006841.full

72. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol. 2020;83(April):104351. Available from: https://doi.org/10.1016/j.meegid.2020.104351