

Review

## On the Origin of SARS-CoV-2: Did Cell Culture Experiments Lead to Increased Virulence of the Progenitor Virus for Humans?

BERND KAINA

*Institute of Toxicology, University Medical Center, Mainz, Germany*

**Abstract.** *We are currently in a rapidly expanding pandemic of the SARS-CoV-2 virus, which originated in the city of Wuhan in central China. The COVID-19 disease is now spread worldwide and has tremendous socio-economic consequences. The origin of the virus can be reconstructed through epidemiological studies and, even more so, from genome comparisons. How the evolution of the virus and the transition to humans might have happened is the subject of much speculation. It is considered certain that the virus is of animal origin and very likely passed from bats to humans in a zoonotic event. An intermediate host was postulated, but many SARS-like bat viruses have the ability to infect human cells directly, which has been shown experimentally by scientists in the Wuhan Institute of Virology using collected specimens containing virus material from horseshoe bats. The propagation of SARS-like bat viruses in cell culture allowed experiments aimed at increasing the infectivity of the virus and adaptation to human cells. This article summarizes the unique properties of SARS-CoV-2 and focusses on a specific sequence encoding the spike protein. Possible scenarios of virus evolution are discussed, with particular emphasis on the hypothesis that the virus could have emerged unintentionally through routine culture or gain-of-function experiments in a laboratory, where it was optimally adapted to human cells and caused cryptic infections among workers who finally spread the virus causing the pandemic.*

This article is freely accessible online.

*Correspondence to:* Bernd Kaina, Institute of Toxicology, University Medical Center, Obere Zahlbacher Str. 67, D-55131 Mainz, Germany. E-mail: kaina@uni-mainz.de

*Key Words:* Corona virus, Covid-19, SARS-CoV-2, gain-of-function experiments, pandemic, bat viruses, review.

The SARS-CoV-2 pandemic is not the first one caused by a coronavirus, but it is undoubtedly the most severe. Already in November 2002, an epidemic emanated from the city of Shenzhen in southern China. Because of the serious symptoms affecting the respiratory tract it was referred to as SARS-CoV (CoV for corona virus). Patient zero was very likely a cook specialized in wild animals. Fortunately, the infection came to an end in May 2004, as declared officially by the World Health Organization (WHO). With 8,096 patients infected and 774 registered deaths (corresponding to a mortality rate of 9.7%), the infection went off lightly. The trigger was a virus from the large family of betacoronaviruses. The natural hosts of these corona viruses are bats (1). It is assumed that bats did not infect humans directly, but that there was an intermediate host for SARS-CoV, the Asian musang (*Paradoxurus hermaphroditus*) from the subfamily of the palm roller. This is a nocturnal, tree-dwelling cat that is eaten and is also kept in farms for the purpose of coffee bean fermentation through digestive enzymes of the animal (1).

Another wave of infections began in 2012, probably originating in Saudi Arabia and again caused by a virus belonging to the beta coronavirus family, albeit from a different subgroup. The disease was severe, characterized by fever, coughing, shortness of breath (respiratory syndrome), and was associated with pneumonia, kidney failure and finally multiple organ failure. The virus that caused it was named MERS-CoV (MERS for Middle East respiratory syndrome). Seriously ill patients infected with MERS-CoV were registered until 2018, with the infection not only occurring in the Arabian Peninsula, but also in South Korea and China (in hospitals, probably through travelers). The focus of the occurrence, however, was the Arabian Peninsula, where local outbreaks occurred repeatedly. By January 2018, the WHO registered 2,143 diagnosed patients, 750 of them died. By February 2020, 2,519 cases had been registered, of which 866 were fatal. This corresponds to a mortality rate of around 35%. Regarding MERS, it is also assumed that there

was an animal reservoir, and since coronaviruses are widespread in bats, bats are also considered to be the primary natural host here. However, it is unlikely that in Arab countries, where bats are not caught for the purpose of consumption, the infection passed directly from them to humans. Rather, an intermediate host was identified, namely dromedaries (Arabian camels) and Asian camels, both carry MERS viruses. Thus, up to 74% of the examined dromedaries were serum-positive. Young dromedaries in particular develop acute illness from MERS-CoV, while the infection is mild in Asian camels. The fact that dromedaries act as intermediate hosts explains the clustered occurrence of the infection in the Arab world. It also explains the interest in vaccinating the animals (1, 2).

### The COVID-19 Pandemic

In view of the recurrence of epidemics caused by zoonoses, in 2018 the WHO warned of the likelihood of a new wave of infections and postulated a disease X. This would soon turn out to be COVID-19. The epidemic started in the Chinese city of Wuhan (about 8 million inhabitants) in the province of Hubei in December 2019 and rapidly became a worldwide pandemic with unpredictable socio-economic consequences and personal sufferings. The first case of illness in Wuhan was reported on December 1, 2019 and the first official case, a patient with pneumonia from Wuhan, was reported to the WHO by Chinese authorities on December 31, 2019 (3,4). More cases followed. A fish and wildlife market in Wuhan, the "Huanan seafood and wildlife market" (hereinafter referred to as Huanan wildlife market), was indicated as the starting point, because around two thirds of the first reported cases worked there or lived near the market (3). Of note, patient zero had no contact to the market (4) and thus it is considered unlikely that the market was the primary infection point. The first 99 patients examined had an average age of 56 years. Two thirds of them were men and half of them suffered from chronic diseases (3). The start of the epidemic is officially given in the first description of the virus as December 12, 2019 in Wuhan (5). The Huanan Fish and Wildlife Market was closed and decontaminated on January 1<sup>st</sup>, 2020.

As early as January 8, 2020, the Center for Health Control and Prevention in Wuhan announced that a new coronavirus had been isolated from a patient with pneumonia and just 2 days later, on January 10, 2020, the first genome sequence of the novel virus was published. This was confirmed on January 11, 2020 by five additional genome analyses (published online) by the Wuhan Institute of Virology (6). This chronological sequence shows how quickly the cause of the initially mysterious lung disease was found by scientists and doctors in Wuhan and how well they were prepared. The results were soon published in prestigious

journals, with the participation of scientists from Shanghai and Beijing, only about six weeks after the first patient had been registered with symptoms that were initially unassignable (5, 7).

The WHO named the disease "coronavirus disease 2019" (COVID-19). As with SARS-1 and MERS, the cause of the COVID-19 pandemic is a coronavirus (8), which was initially called 2019-nCoV and then renamed to SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) (9). The virus is therefore closely related to SARS-CoV, which triggered the 2002-2004 epidemic mentioned earlier in this paper. As of February 24, 2020, 79,331 cases and 2,618 deaths were registered, still predominantly in China. But then the infection quickly spread globally. On May 25, 2020, there were more than 5 million confirmed cases worldwide, most of them (1.6 million) in the USA, and over 345,000 deaths, and 10 month later, on March 8, 2021, more than 117 million cases and 2,6 million deaths were registered globally.

In contrast to other infectious diseases, COVID-19 affects young people far less than the elderly. Infected with symptoms under 19 years of age represent only 2.4% of all cases. It is certain that old people, especially those with previous illnesses and obesity, have a higher risk of mortality. Asymptomatic carriers can spread the infection, and carriers can be contagious 2 to 5 days before the first symptoms present (among others cough, fever, impaired taste and smell). This explains, in addition to the asymptomatic cases, the rapid spread of the virus *via* droplet infection and inhaled aerosols.

The highly critical cases of SARS-CoV-2 infections with a fatal outcome (30-50% of critical cases) are primarily patients with pre-existing conditions such as cardiovascular diseases, high blood pressure, diabetes, chronic respiratory diseases, and cancer. However, fatal courses of illness have also been reported in young patients without previous illnesses. Although the proportion of very severe cases of COVID-19 is lower than in the SARS-CoV epidemic of 2002/2003 and most cases show a symptom-free or mild course, the person-to-person transmission rate, *i.e.* the infectiousness of the virus, is extremely high. This causes the high speed of spread of the virus and thus the high total number of infected and sick people and consequently the high total number of deaths.

Where does the virus come from, which has since become world-wide known simply as the corona virus? To find an answer to this question, one has to know a little about the virus. Therefore, the virus itself and the infection process, especially at the cellular level, will be discussed first.

### The Virus SARS-CoV-2

The SARS virus measures 125 nanometers in diameter and consists of an envelope of lipids and proteins and a single-stranded nucleic acid molecule (29,903 nucleotides) with a

few genes encoding structural proteins and enzymes that serve for its reproduction and distribution (10). As with other corona viruses, influenza viruses (flu pathogens) and rhinoviruses (cold pathogens), the genetic material does not consist of deoxyribonucleic acid (DNA), but of ribonucleic acid (RNA), *i.e.* the form of nucleic acid that is also produced in our cells when the DNA is copied by RNA polymerases and is used for the synthesis of enzymes and other cellular proteins. Compared to other RNA viruses such as HIV, hepatitis C and influenza, the genome of the coronaviruses is relatively large. RNA viruses are not very robust structures despite their lipid virus envelope since RNA can be broken down by specific enzymes, the ribonucleases. These enzymes are omnipresent, including in the sweat of the hands. Outside of our body, in the dry and in the sun, which is particularly harmful to nucleic acids (RNA has an absorption maximum in ultraviolet light at 260 nm, whereby the absorbed energy oxidizes and cross-links the purine and pyrimidine bases), RNA viruses have a short life. Due to these properties, RNA viruses have only one chance to survive by nestling in places where they are protected and by appearing in very large numbers. The main route of transmission is therefore from person to person. This takes place primarily through droplets and aerosols that are released when speaking, coughing, sneezing, singing, etc. and then inhaled (11, 12).

SARS-CoV-2 is genetically relatively stable. Thus, the influenza virus mutates up to 3 times more often than coronaviruses. The reason for this is a "proofreading" mechanism that coronaviruses have. The RNA polymerase checks the RNA produced during virus replication for errors and corrects them. As a result, fewer mutations accumulate in coronaviruses. On the other hand, coronaviruses tend to take up new genetic information through recombination processes. This high recombinogenic ability has implications if we consider the origin of SARS-CoV-2 and, in general, for virus evolution (13).

### How Does the Infection Occur on the Cellular Level?

In order to enter a cell, viruses first need an anchor, a docking mechanism with which they attach to the host cell. This attachment mechanism explains why viruses are host-specific and organ-specific (tropism) and limit themselves to a certain animal species or to humans. In the case of the SARS-CoV-2 virus, the anchor the virus uses for docking is the ACE2 protein that is present on the surface of lung cells, but also on cells of many other organs such as the kidney, intestinal epithelium, the heart and the vascular endothelium (1). Screening studies revealed that mucus-forming goblet cells and the ciliated epithelium in the nose contain the ACE2 protein, thus, it seems reasonable to conclude that the primary infection and virus replication takes place in the

nose. It is also possible that nerve cells are infected, which would explain the impairment of smell and taste. It is also reasonable to assume that inhalation of small amounts of virus results in symptom-free courses, while inhalation of larger amounts of virus allows it to penetrate into the lower areas of the lungs, the alveoli, and infect them, with the known severe courses. Therefore, the amount of virus particles, together with the immune status, is important determining an asymptomatic, mild or severe form of infection.

Once inhaled deeply or *via* the nose as an intermediate station, the virus attaches to the ACE2 protein on the surface of cells with a spike glycoprotein (S-protein) located on its virus envelope and is then internalized (14). How this process works in detail has not yet been fully clarified. It is certain, however, that the binding of the S-protein to the ACE2 protein (also known as the ACE2 receptor) is not sufficient for the infection, but that other cellular proteins are required. One of them is a protein-cleaving enzyme, a protease, that cleaves the S-protein bound to the ACE2 receptor, a process known as "priming". The protease involved has been identified; it is a serine protease called transmembrane protease serine subtype 2 (TMPRSS2) (15). However, inhibiting this protease does not completely prevent the virus from entering, which is why another protease was suspected, which is additionally required or can alternatively take over priming. This turned out to be the protease furin, which cleaves the S-protein of SARS-CoV-2, but not that of SARS-CoV, which caused the wave of infections in 2002/2003 (16). In fact, the S-protein of SARS-CoV-2 has a furin cleavage site that is missing in SARS-CoV (15, 16). This is a section in the protein with the amino acid sequence proline-arginine-arginine-alanine, which is referred to as the PRRA motif or as polybasic cleavage site (PCS) (also referred to as furin cleavage site). The gain of this sequence with the neighboring furin cleavage site is likely responsible for the high infectivity of SARS-CoV-2, because the binding strength to the ACE2 receptor is similar to SARS-CoV, which is less infectious. The protease furin is membrane-bound, but a secreted form has also been reported (17). These could affect the cleavage of the spike protein on the virus envelope on cells that do not express the furin protease, and thus intensify the infection process.

Recently, another cofactor has been identified that enhances the entry of SARS-CoV-2 into ACE2 expressing cells: neuropilin-1 (NRP1) (18, 19). Human embryonic kidney cells (the HEK-293T cell line) do not express detectable ACE2 and NRP1 and display a low level of infection, but upon transfection of ACE2, NRP1 and the protease TMPRSS2, the cells became highly responsive to SARS-CoV-2 infection and produced a high virus titer. This and other cell culture experiments led to the conclusion that NRP1 together with furin and TMPRSS2 are important entry

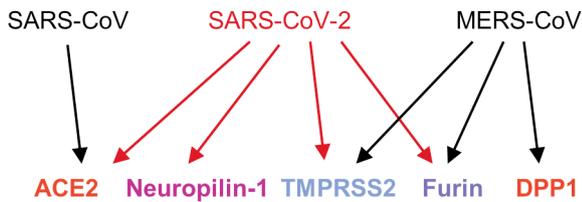


Figure 1. Receptors and proteases required for entrance of SARS-CoV, SARS-CoV-2 and MERS-CoV viruses.

cofactors on the surface of host cells that determine the infection rate by SARS-CoV-2. Interestingly, both TMPRSS2 and NRP1 are abundantly expressed in pulmonary and olfactory cells with highest levels in endothelial cells, which are primarily infected, while ACE2 is expressed only at very low levels, indicating the importance of the cofactors. It is important to note that the MERS virus (MERS-CoV) does not bind to ACE2; it requires a different receptor, namely the dipeptidyl peptidase 4 (DPP4) protein, which is located on the surface of lung cells (20) (Figure 1).

If the ACE2 receptor and the TMPRSS2 and furin proteases are necessary for efficient infection at the cellular level, the question arises in which tissues and cell types these proteins can be found. As already mentioned, ACE2 is mainly expressed in the respiratory tract, but also in the kidneys, the intestine, and the vascular endothelium. Recently published studies have shown that TMPRSS2 and furin are also expressed in cells of the lungs and bronchi (21). It is interesting that there are no gender-specific differences in expression, which is consistent with the fact that no significant differences were found in the infection rates between men and women (22). However, a trend towards an increase in the expression of ACE2 with age has been observed; the expression levels were lower in patients under 50 years of age than in those over 70 years of age. There was no difference in ACE2 expression between smokers and nonsmokers. The levels of expression of ACE2 are generally not high compared to other proteins; among the expressing cells of the respiratory tract it was highest in the pneumocytes protruding into the lumen of the alveoli (21). The expression of ACE2 in the bronchi and in the mucous membrane of the nose may explain early symptoms of the disease, the expression in part of the small intestine, the ileum, the large intestine, the kidney may explain multiple symptoms if these organs are affected. The observation that pericytes of the blood vessels also express the ACE2 protein (21) and thus offer a docking point for SARS-CoV-2 is also very interesting. Pericytes, together with endothelial cells, line the blood vessels and are found in organs with high blood flow, such as the heart. Damage to both the pericytes and the endothelial cells would inevitably lead to necrosis

and thus to vascular damage, with the result of thrombosis and embolism. And this is exactly what can often be observed in severe courses of the infection.

Like ACE2, the protease TMPRSS2 is also found in cells of the nasal mucosa, which supports the assumption that the nose, as the portal of entry, is the primary site of infection (23). ACE2 and TMPRSS2 are also found in the corneal cells of the eye, which supports the notion that the infection can also occur through the eye. The expression of these proteins in children is unclear; however, it is likely that they are also expressed in child tissues, as the infection rates in children are comparable to those in adults, even if the infection is clinically asymptomatic (24).

It should be noted that the replication cycle and the maturation of the virus and its release from the cell are complex processes. Once packaged, the virus is transported to the cell surface and released by exocytosis. However, it has also been observed that after the virus has replicated, it causes the host cell to fuse with the neighboring cells, forming syncytia, and transfers directly into them without being released. Here, too, the spike protein plays a key role (25). The process very likely promotes the spread of the virus in the body and infectivity, as antibodies have no access to the infectious particles and in this way the immune defense would be evaded.

### Where Does SARS-CoV-2 Come From?

This is a very important question that needs to be answered in order to prevent future pandemics. It is considered certain that COVID-19 is a zoonosis, *i.e.* a pathogen that is actually restricted to certain animal species and spread to humans. This is easy to imagine if we remember the mechanism described above by which the virus penetrates human cells. It is reasonable to suppose that the virus also uses the ACE2 receptor (and possibly also NRP1) in bats in order to enter the cell. In the natural host, however, the receptor protein has a slightly different shape than the human protein and better fits S-protein binding so that the infection remains limited to the animal. However, if mutations in the virus RNA result in a change in the structure of the S-protein, this can lead to the virus also attacking cells of other species. Such mutations are very likely to occur over and over again; they are neutral and become part of the virus evolution if the virus has contact to another species such as man.

From the nucleotide sequence of the RNA in the virus genome, one can infer the evolution of SARS-CoV-2 fairly precisely. Thus, it can be safely said that viruses similar to SARS-CoV-2 occur naturally in horseshoe bats (26). More than 60 virus strains have been found in bats, and some species harbor up to 12 different strains (13). The animals are apparently protected against the corona viruses they harbor. Horseshoe bats are found in caves far from Wuhan.

The open question is how did the virus find its way from bats into humans? Where did the decisive zoonotic event take place?

A hypothesis that is still under consideration rests on the fact that about two thirds of the first people known to be sick seemed to have contact with the Huanan wildlife market mentioned above, where live wild animals were also offered and slaughtered. According to press reports, however, no live bats were offered in this market. Since a coronavirus variant with great similarity to SARS-CoV-2 was also found in a pangolin, Chinese researchers proposed that this armadillo could be an intermediate host (27). But how is the virus supposed to get from the bats that catch their food - insects and spiders - in flight at night into the pangolin, whose natural home is Malaysia, and from there into the human lungs? The explanation looks like a hapless assignment of blame. Let's take a closer look at the matter.

### **The Search for the Predecessor of SARS-CoV and SARS-CoV-2**

The SARS-CoV pandemic of 2002/2003 and the MERS-CoV pandemic that started in 2012 have shown that corona viruses pose a significant health threat to humans. From the investigation of the corona viruses causing SARS and MERS scientists have learned that they need cell surface proteins as receptor (ACE2, NRP1 or DPP4) and proteases that mediate the processing of the S-protein. Thus, receptor binding is the crucial bottleneck for transmission across species boundaries (cross-species transmissibility). The early studies also revealed that the S-protein of SARS-CoV is directly involved in binding to the ACE2 receptor (28), and a specific domain in the S-protein is required for this (29). However, up to this point (until 2013) no SARS-like coronaviruses had been found in bats that were able to dock directly onto the human ACE2 receptor.

This changed in 2011/2012 when scientists from the Wuhan Institute of Virology, led by Prof. Zhengli Shi, went on collecting trips. Zhengli Shi heads the Center for Emerging Infectious Diseases at the Wuhan Institute of Virology and is a virologist who is fully committed to bat coronaviruses. In a horseshoe bat colony (of the species *Rhinolophus sinicus*) in a cave in Kunming (in the province of Yunnan, about 1,500 km from Wuhan), she and her group collected 117 throat swabs and fecal samples. Dr. Shi's virologists found what they were looking for in these samples when they amplified and sequenced the virus RNA using RT-nested PCR: 27 of the 117 samples were virus-positive, *i.e.* they contained virus material, and most viruses were from the SARS-like type (named SL-CoV). The samples also contained two new virus sequences (one designated as HCO14 containing approximately 29,000 nucleotides) that entered the literature as strains of coronaviruses that had more than 95% sequence identity at the genome level with SARS-

CoV that caused the 2002/2003 pandemic. This collecting trip brought Zhengli Shi a well-noted publication (30). It is important to note that the bats' throat and feces samples also contained viruses that could be replicated directly in Vero E6 cells (a monkey cell line expressing the ACE2 receptor) and human cells. Thus, a virus clone (RS3367) isolated from the samples used the ACE2 receptor of humans, monkeys, and horseshoe bat cells for cell entry (28). The virus isolated directly from bats consequently has a relatively broad host specificity, which led the researchers to the conclusion (in 2013) that Chinese horseshoe bats represent a natural reservoir of SARS-CoV and that an intermediate host is not necessary to cause an infection in humans: "intermediate hosts may not be necessary for direct human infection by some bat SL-CoVs" (30). In other words, the viruses collected from bats and propagated in cell culture in the Wuhan Institute of Virology had the ability to infect human cells directly. Of course, it has not been investigated whether they can also infect humans themselves, which would not be impossible during intensive contact with bats if students and researchers were not appropriately protected. The viruses potentially pathogenic for humans, brought from the collection trip to Zhengli Shi's laboratory, were propagated in cell lines that express ACE2 naturally or following genetic engineering (upon transfection with the human ACE2 cDNA). When the virus multiplies, the infected cells die and release the virus particles into the cell culture medium. The viruses can now be isolated, further examined and propagated. These are quite normal experimental work steps that have to be carried out under the biosafety level 2 in accordance with the guidelines for biosafety for coronaviruses, and with SARS-CoV under biosafety level 3.

From now on, intensive research into the newly discovered viruses was carried out in Wuhan and other institutions around the world. Together with US scientists (who were even in charge of this), Zhengli Shi's group reported two years later, in 2015, about a genetically engineered chimeric virus that contains the spike protein from one of the bat viruses described above into a non-pathogenic mouse-adapted SARS coronavirus sequence. This chimeric virus proved to be highly pathogenic: it reproduced in human lung cells in cell culture as well as in the mouse lung with the corresponding pathogenesis in animals (31). If the recombinant virus was reisolated after infection, it was still capable of reproduction in the cell culture and in the animal. Available drugs, such as a vaccine against the chimeric virus available in the laboratory, failed in the experiment and the infected mice could not be cured. From these experiments with recombinant viruses that gained a pathogenic function, the authors again drew the conclusion that zoonosis is possible and that the SARS-CoV epidemic of 2002/2003 could be repeated due to viruses circulating in bat populations (31).

Similar investigations followed at the Wuhan Institute for Virology, with further virus strains being isolated from swab and fecal samples from *Rhinolophus sinicus* and other bat species. Although the newly isolated strains had slightly different nucleotide sequences, they all have the gene for the S-protein, which is required to infect human cells (and those of the bat). This has even been shown in a widely used human tumor cell line, the HeLa cells, which expressed human ACE2 after transfection (26). From the comparison of the isolated virus sequence, it could be concluded that in the bat population coronaviruses undergo genetic changes that also affect the spike gene. None of the viruses, however, had properties of human SARS-CoV-2. The fact that, years after the collection campaign in 2011/2012, new viruses could be isolated from the stored fecal samples can be interpreted as an indication that other previously undiscovered sub-strains are stored in the samples at the Wuhan Institute for Virology, possibly also those that have an even stronger sequence similarity to SARS-CoV-2 than those already analyzed.

The institute in Wuhan also examined corona viruses, which originate from bats and cause disease in pigs. For example, in 2016 in the province of Guangdong in China, only about 100 km from the place where the SARS epidemic broke out in 2002/2003, massive pig deaths occurred on a pig farm (a total of 24,000 animals died), with the virus causing an intestinal disease (the severe acute diarrhea syndrome, SADS). The genome of this virus is 98.5% identical to the nucleotide sequence of a virus that the researchers identified in 2016 from samples in a bat cave (of the species *Rhinolophus sinicus*) near the pig farm (32). This sorrowful example of the wave of infections in pigs shows that SARS-CoV and SARS-CoV-2 are not the only zoonotic events that have occurred in connection with bat populations in China. The author is not aware of data indicating that the researchers working with Zhenli Shi or other scientists also tried the potentially human-pathogenic coronaviruses isolated from bats to propagate in pigs, as speculated in the media.

### **How Does Human SARS-CoV-2 Differ from SARS-CoVs Found in Bats and What Makes it Unique?**

The first work in which the nucleotide sequence of SARS-CoV-2 was described and compared with other viruses, was published by Zhengli Shi and coworkers very soon after the outbreak of the wave of infections in Wuhan. It was submitted to Nature on January 20, accepted 9 days later on January 29 and published online on February 3, 2020; it appeared in one of the journal's issues on March 12, 2020 (5). This is a short period of time for an extensive paper with 29 authors from 4 Chinese institutions, starting with the identification of the first patient by the end of December 2019 to virus isolation, sequencing, data acquisition, writing,

approval by all authors, submission, reviewing and revision. In this paper, it is reported that samples from seven patients (six of whom were employed at the animal market) from a hospital in Wuhan were sent to the laboratory. In these samples, coronaviruses were detected that show a high degree of homology to SARS-CoV (80% identity at the nucleotide level and 94% at the amino acid level for some genes). A comparison with other coronaviruses from the institute's collection then showed that the viruses isolated from the patients (now named 2019-nCoV and later renamed SARS-CoV-2) most closely matched a virus strain in the sequence that was previously derived from a bat, the Java horseshoe bat (*Rhinolophus affinis*) that is endemic in the Yunnan Province. This strain, called BatCoV-RaTG13, was found to be 96.2% identical (based on the entire genome) with SARS-CoV-2. The authors concluded that RaTG13 from their coronavirus collection is the closest relative of SARS-CoV-2 and that these viruses differ from other SARS-CoVs in their phylogeny (5).

In the earlier work on coronaviruses from horseshoe bats, which could be propagated directly in human cells *in vitro*, the authors concluded that an intermediate host does not necessarily have to be involved (26, 30). Nevertheless, supported by a publication (27), the Chinese media spread the hypothesis that an intermediate host is the possible carrier and that the Huanan wildlife market in Wuhan is the place where this happened. A pangolin was postulated as the intermediate host, since a SARS-CoV-2-like virus was found in the lungs of a dead animal in the year of the outbreak of the pandemic (33). However, this virus shows a far less degree of identity (91%) in the nucleotide sequence to SARS-CoV-2 than RaTG13 (96%) (34). Therefore, it is reasonable to conclude that the Malaysian pangolin is not the intermediate host and the bat virus RaTG13 or a closely related strain is very likely the immediate predecessor of SARS-CoV-2.

In addition to base-exchange mutations, the genome of SARS-CoV-2 is characterized by a few insertions, including the already mentioned and functionally very important ACE2 receptor binding site, which will be discussed now in more detail. The spike (S) protein of SARS-CoV-2 is about 20-40 nm long and forms a trimeric structure, which serves as a docking point to the ACE2 receptor. As outlined above, the spike protein has also affinity to neuroleptin-1 and is the substrate of at least two membrane-bound proteases, which cleave it to facilitate cellular entry. This complex process appears to be very important in determining the species specificity and organ tropism (35). It should be mentioned that even treatment with trypsin allows infection of human cells with MERS-like viruses, indicating that the proteolytic cleavage of the spike protein together with the receptor binding is decisive for the host specificity (36). Interestingly, the SARS-CoV-2 spike protein does not bind to the ACE2

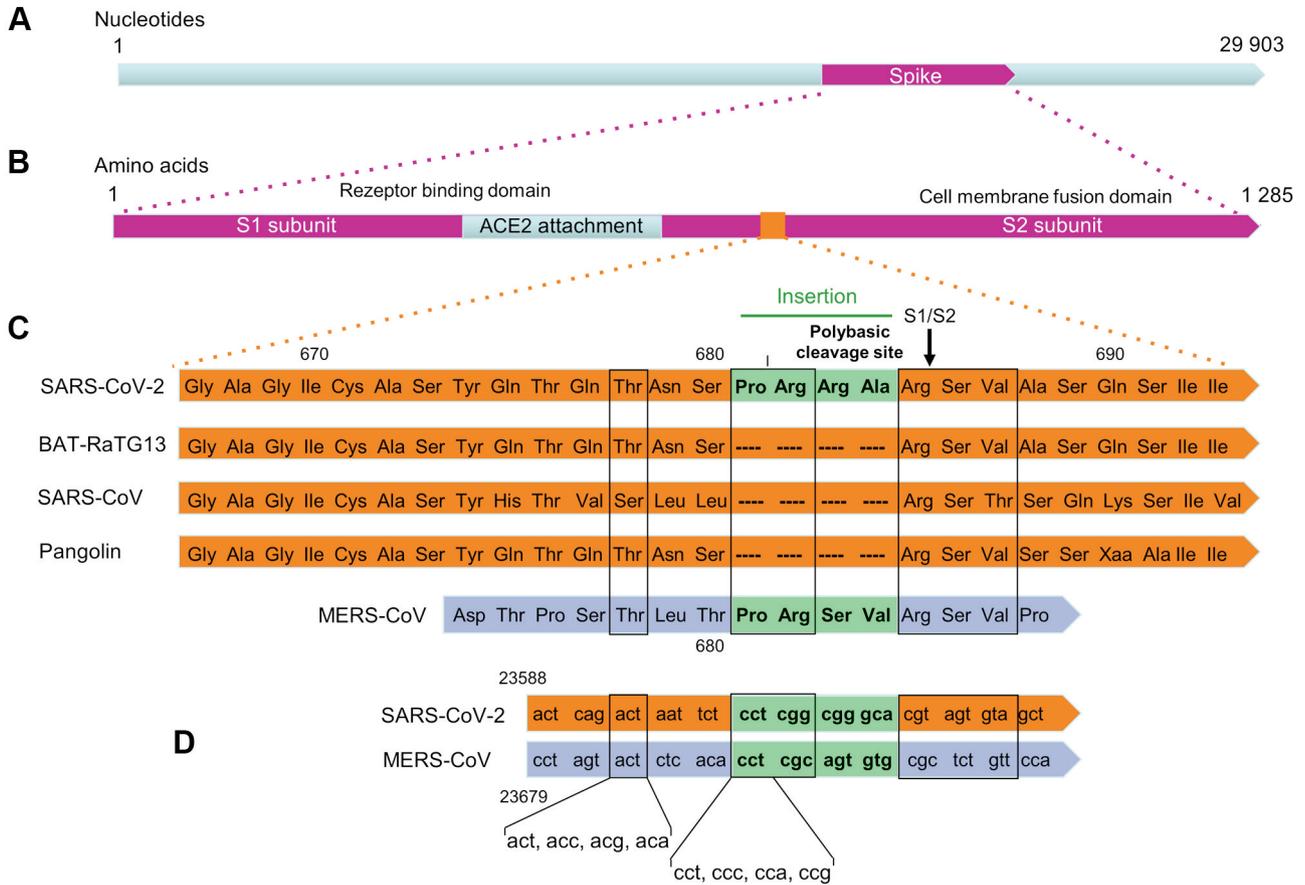


Figure 2. Sequence comparison of the spike protein of SARS-CoV-2 with other coronaviruses whose genome is very similar to that of SARS-CoV-2 (96.2% identity to the Bat-CoV RaTG13 strain isolated from the horseshoe bat; 79.5% identity to SARS-CoV and 91.0% identity to the virus isolated from the pangolin (27)). (A) Genome of SARS-CoV-1. (B) S-protein. After binding to the ACE2 receptor, the spike protein is split into the S1 and S2 subunits. The S1 subunit mediates the binding to ACE2, the S2 subunit the efficient entry into the cell by membrane fusion. It is believed that after binding of the S protein to the ACE2 receptor, two (or even three) cleavage events are necessary for efficient entry into the cell. One is catalyzed by the protease TMPRSS2, the second by furin. (C) The furin cleavage site at the end of the Arg-Arg-Ala-Arg tetrapeptide is marked with an arrow (S1/S2). It is supposed that this cleavage occurs first, followed by cleavage of the S2' site (not shown) through TMPRSS2 (46). The nucleotide sequence that codes for the polybasic cleavage site (see text) rests on an insert that is unique for SARS-CoV-2. A similar insertion can be found in MERS-CoV. The amino acid sequence in the scheme is given using the 3-letter abbreviations in order to make it more understandable for non-biochemists. Modified according to (39) and (16). (D) Sequence comparison of SARS-CoV-2 and MERS-CoV. Sequences are from NCBI database. The possible codon sequences for threonine and proline are given on the bottom of the figure.

receptor more strongly than the spike protein from SL-CoVs, but the furin cleavage of the spike protein results in more efficient entry into human cells, which could explain the high virus infectivity of SARS-CoV-2 (37). The pre-activation of the spike protein by proteases is evidently of utmost importance in the infection process. The spike protein also causes membrane fusion "from without" (38), *i.e.* without entering the cell, which seems to be an additional function gained by SARS-CoV-2 contributing to the high infectiveness.

The ACE2 binding domain of the spike protein is encoded by a sequence region that is considered a hotspot for point mutations (34). A comparison of the nucleotide or amino

acid sequences of the ACE2 receptor binding domain of various SARS coronaviruses revealed that human SARS-CoV-2 differs significantly from other SARS-CoVs in a small sequence, which covers the already mentioned polybasic cleavage site (PCS, also designated as furin cleavage site; Arg, Arg, Ala, Arg; cleavage occurs at the S1/S2 position; Figure 2). This cleavage site was gained by an insertion into the progenitor SARS-CoV-2 genome of a short gene sequence CCTCGGCGGGCA, which encodes four amino acids (Pro-Arg -Arg-Ala, highlighted in green in Figure 2) (39). It results in the aforementioned PCS, which is essential for the entry of the virus into human cells as it

allows the priming after S-protein receptor binding. In bats, in which the docking also takes place *via* the ACE2 receptor, this cleavage is obviously not necessary. Thus, it is evident that human SARS-CoV-2 received this sequence from an unknown source, through recombination with cellular or viral RNA or through multiple spontaneous nucleotide insertions and substitutions, because it cannot be found in any of the previously known SARS coronaviruses, not even in the pangolin (Figure 2). It stands to reason that the acquisition of the insertion making up the PCS/furin cleavage site is closely related to the primary zoonotic event. Thus, the question of how the insertion of the sequence CCTCGGCGGGCA came about is the focus of interest.

It is important to note that all SARS-like coronaviruses known so far do not harbor this insert, while MERS-CoV, which docks to the DPP4 receptor, has an insert in this position of the protein consisting of four amino acids (Pro-Arg-Ser-Val) (Figure 2C). Furin cleavage sites are also found in other viruses in attachment proteins, including HIV, where the protease plays a role in entering the cell. However, a close similarity was found with the sequence of MERS-CoV. In Figure 2D, the nucleotide sequence around the PCS is compared between SARS-CoV-2 and MERS-CoV. Interestingly, on position 678 threonine is encoded by the same triplet ACT in SARS-CoV-2 and MERS-CoV, and on position 681 proline is encoded again by the same triplet CCT and on position 682 arginine by CGG and CGC. Thus, there is not only a strong identity on the amino acid level, but also on the nucleotide level between SARS-CoV-2 and MERS-CoV. Given the code redundancy, the probability for an identical sequence encoding threonine (pos 678; which can be encoded by ACA, ACG, ACC and ACT) is  $0.25 \times 0.25 = 0.0625$ , and that of proline (pos 681; encoded by CCT, CCC, CCA and CCG) is again 0.0625. The overall probability harboring the same nucleotide sequence in these two positions is 0.0039. In position 682 with arginine coded by the sequence CGG (SARS-CoV-2) and CGC (MERS-CoV) we are again faced with a coincidence (nucleotides CG) of low probability [arginine is encoded by the codons CG (G,C,A,T), AGG, AGA]. In conclusion, there is a remarkable identity on amino acid and nucleotide level in and around the PCS between SARS-CoV-2 and MERS-CoV. This supports the hypothesis that the PCS/furin cleavage site was gained by a recombination event(s) involving these virus sequences. This notion is important in considering possible zoonotic events, placing laboratory events in the realm of the highly possible.

In this context, it is important to note that a sequence comparison of SARS-CoV-2 with other viruses revealed a 117-nucleotide sequence in the virus genome that is 94.6% identical to a human intron sequence of the netrin G1 gene. Several other viruses also contain human sequences, but they are much shorter (*e.g.*, SARS-CoV contains a 41-nucleotide sequence). MERS-CoV does not contain a human sequence (40). The

presence of a human sequence in SARS-CoV-2 supports the hypothesis that the progenitor was propagated in human cells where it gained the sequence by a recombination event.

### What Scenarios are Conceivable for Zoonosis?

Zoonosis can develop in different scenarios in the natural environment. First of all, one could speculate that the virus with the complete sequence of SARS-CoV-2 (including the furin cleavage site) is also present in bats, but has not yet been discovered. It is also conceivable that RaTG13 or another predecessor of SARS-CoV-2 with a similar or higher level of identity than RaTG13 directly infected human individuals, but this pathogen had a very weak virulence and therefore the infection initially went unnoticed.

In view of the discussion of a virus evolution before zoonosis took place, the possibility has been favoured that selection of a bat virus similar to RaTG13 has occurred in another animal that is equipped with an ACE2 receptor protein (and neuropilin-1) similar to that of humans. The selection in an intermediate host resulted in the creation of a spike protein with a receptor binding domain and protease preactivation sites, which binds more efficiently than RaTG13 to the human ACE2 protein and enters the cell upon processing as described above. It was proposed that this happened in the Malaysian pangolin, in which coronaviruses were found with similarity to RaTG13 (31). However, this assumption rests on a single-case report and there was no systematic search for SL-CoVs in this species. Furthermore, the coronavirus found in the pangolin did not harbor the PCS/furin cleavage site that is typical for SARS-CoV-2 (Figure 2). Also, there is no explanation as to how the virus in an intermediate host can gain the PCS. Although insertions and deletions occur frequently in corona viruses (33), there must have been a selection pressure in order to favor the existence of this cleavage site. It follows that the most likely intermediate host is a species whose ACE2 and neuropilin-1 receptors are similar to those of humans, including the processing mechanism. In addition, one must assume that it is a species with a high population density in order to make recombination events and natural selection efficient through frequent transmission of the virus and that it has frequent contact with humans. Pangolins have a low rate of reproduction, which would probably have been zero at the Huanan wildlife market in Wuhan, if they were offered there. As a selection medium for adaptation to humans, pangolins (or similar exotic animals) are therefore not likely to be a particularly good choice for the virus.

In view of this, primary infection from animals on the wildlife market is unlikely to have happened. One could also speculate that house cats from families of asymptomatic infected people were the intermediate host. This would be even more likely since cats are equipped with an ACE2

protein that is very similar to humans (41), and house cats (as well as big cats in zoos) can become infected with SARS-CoV-2. Thus, the virus can pass from humans to cats and among cats themselves (42). If it is proven that house cats can also infect humans, it is reasonable to assume that the spatial proximity between house cats and humans could have favored a virus evolution. All these scenarios, however, do not provide an answer on how a bat virus found the way into an intermediate host, be it the pangolin, cat or others.

Another possibility to be considered is that there was no intermediate animal host, but natural selection took place in humans directly. The predecessor of SARS-CoV-2 would therefore have jumped directly from the bat into humans (*e.g.* by inhaling dust or droplets/aerosols exhaled by bats, from which SL-CoVs were isolated, or when hunting and preparing the animals for consumption) and initially there would have been an undetected human-to-human transmission. The virulence was initially very low, but increased over time when the furin cleavage site and other supportive mutations were gained. In this scenario, selection took place in humans until the insertion of the furin cleavage site was perfect. Thus, the virus gained the property to be efficiently propagated in humans and clusters of infection were formed due to high infectivity, which allowed the virus to survive evolutionarily. This scenario assumes that there was a period of undetected infection and transmission even before the furin cleavage site was incorporated into the viral genome. It has been learned from MERS-CoV that human diseases can be caused by the corona virus jumping from the dromedary to humans, resulting in permanent transmission and reproduction in humans without previous adaptation (43). However, this scenario could occur everywhere where people are in close contact to bats or another primary host. It does not explain the origin in Wuhan. Also, the scenario does not answer the question of the origin of the PCS/furin cleavage site, whose gain was obviously a "clonal" event.

A hypothesis intensively discussed in scientific and public media is that SARS-CoV-2 is "man-made", *i.e.* it represents a laboratory construct or was purposefully manipulated. Thus, it is assumed that the virus originated from the Wuhan Institute of Virology or the Municipal Institute for Disease Control, which is located in the immediate vicinity of the Huanan wildlife market. The speculation "man-made" has three aspects that will be discussed: a) the intentional construction through genetic manipulation, b) the intentional selection in the laboratory for high infectivity *in vitro* and in the test animal, and c) the accidental evolution and human adaptation of the virus in the laboratory.

As outlined above, SARS coronaviruses were genetically engineered in several laboratories, including in Wuhan, and chimeric viruses were produced that contained nucleotide sequences from different virus strains (31, 44, 45). But the human SARS-CoV-2 shows no evidence of this type of

genetic manipulation; it does not bear signs of gross genetic changes (96.2% identity to RaTG13) and, therefore, does not appear to be a simple fusion product of different viruses. The changes that make it different from the putative bat progenitor are more subtle. They are technically feasible, but this is above the scope of this review. It is more difficult to assess whether options (b) or (c) apply. These scenarios also postulate human involvement and are based on the fact that at the place of origin of the pandemic, in Wuhan, intensive work was and is being carried out on SARS-CoV, involving human cell infection, large-scale virus propagation and experimenting with them *in vitro* and in experimental animals.

As already mentioned, the virus propagation usually takes place in cultured human cells, whereby, among others, primary epithelial cells of the lung (such as Calu-3), primary cells of the kidney and established lines transfected with ACE2 (Huh7, HCT18, HeLa) were used. The Vero-E6 line, which comes from the green monkey, is also popular. It is conceivable that under these conditions of *in vitro* propagation, the virus from bats had sufficient time to adapt to the human ACE2 (and neuropilin1) receptor and to optimize its reproduction in human cells. It cannot be excluded that the SARS-CoV-2 specific furin cleavage site was acquired spontaneously during virus replication in the cell culture, because some human proteins also bear a furin cleavage site and, therefore, the corresponding coding mRNA could be a natural reservoir for the insert.

It is also conceivable that when cells in culture were coinfecting with the predecessor of SARS-CoV-2 and another virus strain that contains the PCS/furin cleavage site, the sequence was transferred to the predecessor virus as a result of a recombination event. Interestingly, MERS-CoV contains a furin-specific cleavage site, which is very similar to the one found in SARS-CoV-2. As outlined above, there are 6 identical amino acids between position 678 and 687 in SARS-CoV-2 and MERS-CoV. Moreover, the nucleotide sequence encoding the identical amino acids is very similar, which is even more than surprising in view of codon degeneracy (Figure 2D). Gain of function through the PCS could have happened unintentionally, *e.g.* after introducing different viruses into a cell to see whether they complement each other functionally ("in trans"), or purposefully in the context of gain-of-function experiments. In both cases, propagation of the progenitor of SARS-CoV-2 (*e.g.* RaTG13) together with MERS-CoV is likely to lead to the selection of a virus that gains functions of both viruses, thus increasing its infection rate *in vitro*. The cells used were equipped with the ACE2 receptor. If these cells also harbored the neuropilin-1 membran protein, selection for a virus showing an even higher virus titer upon propagation *in vitro* is conceivable. It should be noted that cell lines harboring ACE2 together with NRPI and other desired properties can be generated by routine techniques (transient or stable transfection, lentivirus transduction) and have actually been used for experimental purposes (18, 19).

According to the scenario above, it is possible that primary zoonosis occurred unintentionally while working with cell cultures and cell culture supernatants that contain coronaviruses (*e.g.* through inhalation of aerosols or through smear infections, *e.g.* through improper disposal of biowaste). The selection required for the high infectivity could have occurred during routine cell culture and virus propagation, during selection for high infectivity and could have continued after the cryptic infection of the human cell-adapted virus to humans (*e.g.* laboratory workers). As already mentioned, work with coronaviruses in the laboratory can be carried out under low safety (level 2) conditions, with SARS coronaviruses under intermediary safety (level 3) conditions and genetic engineering of viruses under highest safety (level 4) conditions. The author is not aware of the conditions under which the collections in the bat caves and the processing and propagation of the original samples from 2011/2012 took place. The Wuhan Institute of Virology has a P4 laboratory, the highest level of safety and the only laboratory of its kind in China. However, it was not operational until 2015. Safety level 4 is also not mandatory for working with SARS viruses. Therefore, contamination during work or during disposal of the biowaste cannot be ruled out, especially since the primary infections with viruses without a perfect PCS are likely to have occurred latently and without any signs of disease. The scenario described here ultimately assumes that cryptic infections occurred unnoticed during the experimental work. These laboratory events do not necessarily require the assumption of a "laboratory accident", as we do not see it as an accident when a doctor in a hospital becomes infected with the virus. Nevertheless, an accident cannot be excluded, *e.g.* by improper work on a clean bench, during supernatant centrifugation, virus enrichment or during decontamination of laboratory waste by autoclaving. According to this scenario, it is not necessary to assume an intentional manipulation of the predecessor virus. Although genetic manipulations required for purposeful changing the nucleotide sequence of the virus are technically possible, the virus evolution could happen during long-term cultivation in human cells without their means.

### Comparative Assessment

The SARS-CoV-2 coronavirus exhibits some unusual properties that need to be considered in substantiating the laboratory hypothesis (Figure 3):

a) The high level of infectivity, the low proportion of infected individuals that became ill, and the low level of lethality (ratio of deceased to ill patients). Thus, the lethality for other corona virus infections is high (9.6% for SARS/2002; 34.4% for MERS/2012; 40.4% for Ebola; 80.0% for Marburg) compared to COVID-19 (2.1%). High

infectivity and symptom-free carriers are expected to favor the spread of infection if a laboratory event happened. The low disease and lethality rate enables the virus to propagate in a cryptic, undetected way in humans.

b) Bat RaTG13 and other SL-CoVs are able to infect human cells directly. The insertion of 12 nucleotides in the spike protein sequence turned the bat progenitor virus (RaTG13 or a similar strain) to a more aggressively growing virus characterized by high ACE2 binding and optimised RBD processing by cellular membrane-bound proteases. Thus, the gain of the PCS/furin cleavage site enhanced the infection rate in human cells.

c) The spike insert sequence of SARS-CoV-2 is present in MERS, with 50% identity on amino acid level and some identity on nucleotide level, which is highly unlikely to be gained accidentally given the degeneration of the code (Figure 2D). It is conceivable that coinfection of human cells expressing ACE2 and DPP4 with RaTG13 (or another bat progenitor) with MERS-CoV led to the selection of a hybrid virus that gained through recombination the insert and the furin cleavage site. This is a reasonable process since coronaviruses are highly recombinogenic. As a result, virus propagation would be enhanced *in vitro*. Thus, without selection pressure, cotransduction experiments with human cells *in vitro* would lead to a virus strain with improved properties regarding cell entrance and propagation.

d) The presence of a human sequence in the virus genome (40) strongly indicates that SARS-CoV-2 was propagated in human cells before it caused the pandemic.

e) SARS-CoV-2 infection is strongly enhanced by supportive factors such as neuropilin-1 and the proteases MPRSS2 and furin. It is highly unlikely that this complex scenario that facilitates the virus entrance was gained in a single step and in an intermediate animal host. It is more likely that these properties were obtained during virus propagation in human cells, which were engineered (ACE2, DPP4) or in transient transfection experiments in order to improve the conditions for virus entrance. Overall, SARS-CoV-2 appears to be ideally adapted to human cells, which supports the laboratory hypothesis (Figure 3).

It should be emphasized again that the laboratory hypothesis does not posit that SARS-CoV-2 was genetically engineered on purpose, in simple words "a laboratory construct". The hypothesis rather states that SARS-CoV-2 is an unintended byproduct of gain-of-function and cotransfection/cotransduction experiments using human (genetically engineered) cell lines *in vitro*. Selection occurred during virus propagation for a bat virus that is best equipped with tools using supportive factors of human cells and therefore best adapted to humans. There are many conceivable scenarios how transmission could occur in the laboratory, *e.g.* through aerosols during the work or during handling of waste. Such laboratory events could have occurred repeatedly long before December 2019. They remained undetected because of the cryptic propagation of the virus

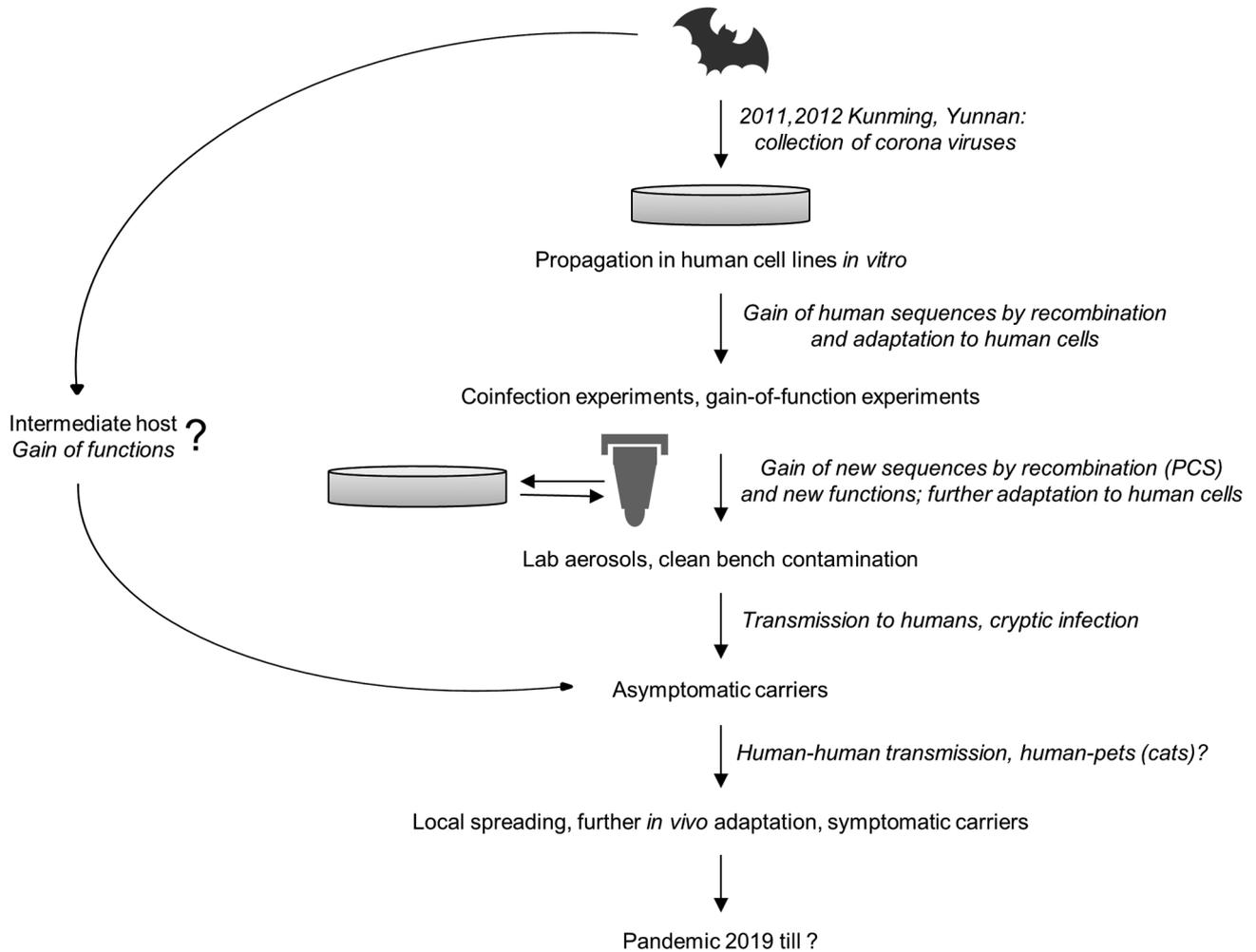


Figure 3. Scenario according to which SARS-CoV-2 could have been evolved. The upper steps in the lab scenario of virus collection and experimentation are well documented in the literature.

especially in young people. Nevertheless, a laboratory accident cannot be excluded. In contrast to this, the intermediate host hypothesis rests on many more assumptions. Thus, it is unclear how the virus from nocturnal bats found the way into the intermediate host and how the selection for optimal human propagation could take place there. The intermediate host hypothesis is therefore regarded as less likely (Figure 3).

### Summary and Conclusion

The current COVID-19 pandemic caused by the SARS-CoV-2 coronavirus has already caused immeasurable suffering and economic damage, and its long-term socio-economic impact cannot yet be assessed. Never before has an event caused more death and long-term suffering, and the associated crisis management brought trade, traffic, travel, social activities and

even family contacts to a standstill. Therefore, the question of how this pandemic came about and when the primary zoonotic event took place is important to answer in order to be able to prevent further zoonotic events of this kind.

Although it was initially believed that the virus could not be transmitted from person to person, the opposite soon turned out. The intensity of the wave of infection is favored by the high infectivity of the virus, by symptom-free carriers and, in symptomatic cases, by a symptom-free period lasting several days. It is considered as certain that SARS-CoV-2 is of bat origin and that there must have been gain of function events through insertions and point mutations by which adaptation of the predecessor virus occurred to human cells equipped with the ACE2 receptor, with neuropilin-1 and the proteases furin and TMPRSS2, all of which are ideally suited for an optimal virus entry into human cells.

SARS-CoV-2 has the greatest similarity in the nucleic acid and amino acid sequence to a virus (RaTG13) that has been isolated from swabs of bats, endemic in central China, at the Institute of Virology in Wuhan and propagated there in cell culture. Although critical sequences in the spike protein, in particular the insertion of a twelve-nucleotide section in the polybasic cleavage site, could have been acquired naturally, for example through recombination events with a SARS-like corona virus harboring the same sequence, it is highly unlikely that the selection occurred in bats or in an intermediate animal host. However, a scenario seems likely, according to which the selection for a highly infectious agent took place in human cells, notably in cell culture, as a byproduct of virus propagation and experimental work. Sequence comparison revealed that the insert creating the PCS/furine cleavage site is partially identical to an insert in the spike gene of MERS-CoV. Based on this, the hypothesis was proposed that during coinfection of human cells equipped with ACE2 and DPP4, RaTG13 (or a similar progenitor) gained the sequence from MERS-CoV (or a similar virus harboring the sequence) by recombination, through which it became better propagating *in vitro*, which is beneficial for the experimentation (selection for high virus titer). At the same time the infectivity of the virus was enhanced allowing unintentional, cryptic infections of employees, which was the starting point for the pandemic.

Furthermore, the possibility that the progenitor virus was propagated in human ACE2 transgenic animals, which would accelerate human adaptation, should also be considered. At this point it should be remembered that releases of pathogenic viruses from the laboratory have already happened, which was a matter of serious concern in China years ago. Back in 2003, in Kunming (province Yunnan), a hantavirus outbreak occurred, and it was attributed to laboratory rats in which two hantavirus strains multiplied and generated a new type of virus through recombination. Students who had worked with the animals became infected with this. The authors of the report (published in 2010) concluded prophetically that "This study sends a timely warning that laboratory exposure remains an important source of hantavirus infection in China and that new strains continue to emerge via reassortment and recombination of the RNA genome segments" (47). There is no reason to believe that the same cannot happen with other virus strains. The coincidence of the outbreak of the disease COVID-19 in Wuhan and intensive ongoing work with SARS-like viruses at a research institute located there, housing the largest corona virus bank in Asia, has also sparked the public discussion whether gain-of-function experiments contributed to the pandemic. Even though evidence is lacking that the virus was intentionally genetically engineered, the possibility that the virus could have emerged unintentionally through laboratory experiments should lead to a rethinking of the need for gain-of-function

experiments aimed at enhancing the pathogenicity of a disease-causing agent.

## Conflicts of Interest

The Author declares that there are no conflicts of interest related to this work.

## References

- de Wit E, van Doremalen N, Falzarano D and Munster VJ: SARS and MERS: Recent insights into emerging coronaviruses. *Nat Rev Microbiol* 14(8): 523-534, 2016. PMID: 27344959. DOI: 10.1038/nrmicro.2016.81
- Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD and Fouchier RA: Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367(19): 1814-1820, 2012. PMID: 23075143. DOI: 10.1056/NEJMoa1211721
- Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X and Zhang L: Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 395(10223): 507-513, 2020. PMID: 32007143. DOI: 10.1016/S0140-6736(20)30211-7
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J and Cao B: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395(10223): 497-506, 2020. PMID: 31986264. DOI: 10.1016/S0140-6736(20)30183-5
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF and Shi ZL: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579(7798): 270-273, 2020. PMID: 32015507. DOI: 10.1038/s41586-020-2012-7
- Gralinski LE and Menachery VD: Return of the Coronavirus: 2019-nCoV. *Viruses* 12(2): 135, 2020. PMID: 31991541. DOI: 10.3390/v12020135
- Jiang S and Shi ZL: The first disease X is caused by a highly transmissible acute respiratory syndrome coronavirus. *Virology* 535(3): 263-265, 2020. PMID: 32060789. DOI: 10.1007/s12250-020-00206-5
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W and China Novel Coronavirus Investigating and Research Team: A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382(8): 727-733, 2020. PMID: 31978945. DOI: 10.1056/NEJMoa2001017
- Simmonds P, Gorbalenya AE, Harvala H, Hovi T, Knowles NJ, Lindberg AM, Oberste MS, Palmenberg AC, Reuter G, Skern T, Tapparel C, Wolthers KC, Woo PCY and Zell R: Recommendations for the nomenclature of enteroviruses and rhinoviruses. *Arch Virol* 165(3): 793-797, 2020. PMID: 31980941. DOI: 10.1007/s00705-019-04520-6
- Khailany RA, Safdar M and Ozaslan M: Genomic characterization of a novel SARS-CoV-2. *Gene Rep* 19: 100682, 2020. PMID: 32300673. DOI: 10.1016/j.genrep.2020.100682

- 11 Stadnytskyi V, Bax CE, Bax A and Anfinrud P: The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proc Natl Acad Sci USA* 117(22): 11875-11877, 2020. PMID: 32404416. DOI: 10.1073/pnas.2006874117
- 12 van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, Tamin A, Harcourt JL, Thornburg NJ, Gerber SI, Lloyd-Smith JO, de Wit E and Munster VJ: Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N Engl J Med* 382(16): 1564-1567, 2020. PMID: 32182409. DOI: 10.1056/NEJMc2004973
- 13 Cyranoski D: Profile of a killer: The complex biology powering the coronavirus pandemic. *Nature* 581(7806): 22-26, 2020. PMID: 32367025. DOI: 10.1038/d41586-020-01315-7
- 14 Gallagher TM and Buchmeier MJ: Coronavirus spike proteins in viral entry and pathogenesis. *Virology* 279(2): 371-374, 2001. PMID: 11162792. DOI: 10.1006/viro.2000.0757
- 15 Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C and Pöhlmann S: SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181(2): 271-280.e8, 2020. PMID: 32142651. DOI: 10.1016/j.cell.2020.02.052
- 16 Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG and 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* 176: 104742, 2020. PMID: 32057769. DOI: 10.1016/j.antiviral.2020.104742
- 17 Vidricaire G, Denault JB and Leduc R: Characterization of a secreted form of human furin endoprotease. *Biochem Biophys Res Commun* 195(2): 1011-1018, 1993. PMID: 8373381. DOI: 10.1006/bbrc.1993.2145
- 18 Daly JL, Simonetti B, Klein K, Chen KE, Williamson MK, Antón-Plágaro C, Shoemark DK, Simón-Gracia L, Bauer M, Hollandi R, Greber UF, Horvath P, Sessions RB, Helenius A, Hiscox JA, Teesalu T, Matthews DA, Davidson AD, Collins BM, Cullen PJ and Yamauchi Y: Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* 370(6518): 861-865, 2020. PMID: 33082294. DOI: 10.1126/science.abd3072
- 19 Cantuti-Castelvetri L, Ojha R, Pedro LD, Djannatian M, Franz J, Kuivanen S, van der Meer F, Kallio K, Kaya T, Anastasina M, Smura T, Levanov L, Szirovicza L, Tobí A, Kallio-Kokko H, Österlund P, Joensuu M, Meunier FA, Butcher SJ, Winkler MS, Mollenhauer B, Helenius A, Gokce O, Teesalu T, Hepojoki J, Vapalahti O, Stadelmann C, Balistreri G and Simons M: Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* 370(6518): 856-860, 2020. PMID: 33082293. DOI: 10.1126/science.abd2985
- 20 Raj VS, Mou H, Smits SL, Dekkers DH, Müller MA, Dijkman R, Muth D, Demmers JA, Zaki A, Fouchier RA, Thiel V, Drosten C, Rottier PJ, Osterhaus AD, Bosch BJ and Haagmans BL: Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495(7440): 251-254, 2013. PMID: 23486063. DOI: 10.1038/nature12005
- 21 Lukassen S, Chua RL, Trefzer T, Kahn NC, Schneider MA, Muley T, Winter H, Meister M, Veith C, Boots AW, Hennig BP, Kreuter M, Conrad C and Eils R: SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J* 39(10): e105114, 2020. PMID: 32246845. DOI: 10.15252/embj.20105114
- 22 Brüßow H: The novel coronavirus - A snapshot of current knowledge. *Microb Biotechnol* 13(3): 607-612, 2020. PMID: 32144890. DOI: 10.1111/1751-7915.13557
- 23 Sungnak W, Huang N, Bécavin C, Berg M, Queen R, Litvinukova M, Talavera-López C, Maatz H, Reichart D, Sampaziotis F, Worlock KB, Yoshida M, Barnes JL and HCA Lung Biological Network: SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med* 26(5): 681-687, 2020. PMID: 32327758. DOI: 10.1038/s41591-020-0868-6
- 24 Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, Liu X, Wei L, Truelove SA, Zhang T, Gao W, Cheng C, Tang X, Wu X, Wu Y, Sun B, Huang S, Sun Y, Zhang J, Ma T, Lessler J and Feng T: Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *Lancet Infect Dis* 20(8): 911-919, 2020. PMID: 32353347. DOI: 10.1016/S1473-3099(20)30287-5
- 25 Fehr AR and Perlman S: Coronaviruses: An overview of their replication and pathogenesis. *Methods Mol Biol* 1282: 1-23, 2015. PMID: 25720466. DOI: 10.1007/978-1-4939-2438-7\_1
- 26 Yang XL, Hu B, Wang B, Wang MN, Zhang Q, Zhang W, Wu LJ, Ge XY, Zhang YZ, Daszak P, Wang LF and Shi ZL: Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. *J Virol* 90(6): 3253-3256, 2015. PMID: 26719272. DOI: 10.1128/JVI.02582-15
- 27 Zhang T, Wu Q and Zhang Z: Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr Biol* 30(7): 1346-1351.e2, 2020. PMID: 32197085. DOI: 10.1016/j.cub.2020.03.022
- 28 Li W, Moore MJ, Vasiliouva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H and Farzan M: Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426(6965): 450-454, 2003. PMID: 14647384. DOI: 10.1038/nature02145
- 29 Wong SK, Li W, Moore MJ, Choe H and Farzan M: A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* 279(5): 3197-3201, 2004. PMID: 14670965. DOI: 10.1074/jbc.C300520200
- 30 Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang YJ, Luo CM, Tan B, Wang N, Zhu Y, Cramer G, Zhang SY, Wang LF, Daszak P and Shi ZL: Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 503(7477): 535-538, 2013. PMID: 24172901. DOI: 10.1038/nature12711
- 31 Menachery VD, Yount BL Jr, Debbink K, Agnihothram S, Gralinski LE, Plante JA, Graham RL, Scobey T, Ge XY, Donaldson EF, Randell SH, Lanzavecchia A, Marasco WA, Shi ZL and Baric RS: A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* 21(12): 1508-1513, 2015. PMID: 26552008. DOI: 10.1038/nm.3985
- 32 Zhou P, Fan H, Lan T, Yang XL, Shi WF, Zhang W, Zhu Y, Zhang YW, Xie QM, Mani S, Zheng XS, Li B, Li JM, Guo H, Pei GQ, An XP, Chen JW, Zhou L, Mai KJ, Wu ZX, Li D, Anderson DE, Zhang LB, Li SY, Mi ZQ, He TT, Cong F, Guo PJ, Huang R, Luo Y, Liu XL, Chen J, Huang Y, Sun Q, Zhang XL, Wang YY, Xing SZ, Chen YS, Sun Y, Li J, Daszak P, Wang LF, Shi ZL, Tong YG and Ma JY: Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin.

- Nature 556(7700): 255-258, 2018. PMID: 29618817. DOI: 10.1038/s41586-018-0010-9
- 33 Liu P, Chen W and Chen JP: Viral metagenomics revealed sendai virus and coronavirus infection of Malayan pangolins (*Manis javanica*). *Viruses* 11(11): 979, 2019. PMID: 31652964. DOI: 10.3390/v11110979
- 34 Zhang YZ and Holmes EC: A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell* 181(2): 223-227, 2020. PMID: 32220310. DOI: 10.1016/j.cell.2020.03.035
- 35 Pillay TS: Gene of the month: The 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. *J Clin Pathol* 73(7): 366-369, 2020. PMID: 32376714. DOI: 10.1136/jclinpath-2020-206658
- 36 Menachery VD, Dinnon KH 3rd, Yount BL Jr, McAnarney ET, Gralinski LE, Hale A, Graham RL, Scobey T, Anthony SJ, Wang L, Graham B, Randell SH, Lipkin WI and Baric RS: Trypsin treatment unlocks barrier for zoonotic bat coronavirus infection. *J Virol* 94(5): 2020. PMID: 31801868. DOI: 10.1128/JVI.01774-19
- 37 Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A and Li F: Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci USA* 117(21): 11727-11734, 2020. PMID: 32376634. DOI: 10.1073/pnas.2003138117
- 38 Theuerkauf SA, Michels A, Riechert V, Maier TJ, Flory E, Cichutek K and Buchholz CJ: Quantitative assays reveal cell fusion at minimal levels of SARS-CoV-2 spike protein and fusion from without. *iScience* 24(3): 102170, 2021. PMID: 33585805. DOI: 10.1016/j.isci.2021.102170
- 39 Andersen KG, Rambaut A, Lipkin WI, Holmes EC and Garry RF: The proximal origin of SARS-CoV-2. *Nat Med* 26(4): 450-452, 2020. PMID: 32284615. DOI: 10.1038/s41591-020-0820-9
- 40 Lehrer S and Rheinstein PH: Human gene sequences in SARS-CoV-2 and other viruses. *In Vivo* 34(3 Suppl): 1633-1636, 2020. PMID: 32503822. DOI: 10.21873/invivo.11954
- 41 Wan Y, Shang J, Graham R, Baric RS and Li F: Receptor recognition by the novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS coronavirus. *J Virol* 94(7): 2020. PMID: 31996437. DOI: 10.1128/JVI.00127-20
- 42 Halfmann PJ, Hatta M, Chiba S, Maemura T, Fan S, Takeda M, Kinoshita N, Hattori SI, Sakai-Tagawa Y, Iwatsuki-Horimoto K, Imai M and Kawaoka Y: Transmission of SARS-CoV-2 in domestic cats. *N Engl J Med* 383(6): 592-594, 2020. PMID: 32402157. DOI: 10.1056/NEJMc2013400
- 43 Dudas G, Carvalho LM, Rambaut A and Bedford T: MERS-CoV spillover at the camel-human interface. *Elife* 7: 2018. PMID: 29336306. DOI: 10.7554/eLife.31257
- 44 Ren W, Qu X, Li W, Han Z, Yu M, Zhou P, Zhang SY, Wang LF, Deng H and Shi Z: Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *J Virol* 82(4): 1899-1907, 2008. PMID: 18077725. DOI: 10.1128/JVI.01085-07
- 45 Hou Y, Peng C, Yu M, Li Y, Han Z, Li F, Wang LF and Shi Z: Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. *Arch Virol* 155(10): 1563-1569, 2010. PMID: 20567988. DOI: 10.1007/s00705-010-0729-6
- 46 Bestle D, Heindl MR, Limburg H, Lam TV, Pilgram O, Moulton H, Stein DA, Hards K, Eickmann, M, Dolnik O, Rohde C, Klenk, HD, Garten W, Steinmetzer, T and Böttcher-Friebertshäuser E: TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci Alliance* 3(9): e202000786, 2020. PMID: 32703818. DOI: 10.26508/lsa.202000786
- 47 Zhang Y, Zhang, H, Dong X, Yuan J, Zhang H, Yang, X, Zhou P, Ge X, Li Y, Wang LF and Shi Z: Hantavirus outbreak associated with laboratory rats in Yunnan, China. *Infect Genet Evol* 10: 638-644, 2010. PMID: 20380897. DOI: 10.1016/j.meegid.2010.03.015

Received March 16, 2021  
 Revised March 24, 2021  
 Accepted March 25, 2021