

# Mutations and Epidemiology of SARS-CoV-2 Compared to Selected Corona Viruses during the First Six Months of the COVID-19 Pandemic: A Review

Miriam M. Nzivo<sup>1\*</sup>  and Nancy L.M. Budambula<sup>2</sup> 

<sup>1</sup>School of Biological Sciences, Jomo Kenyatta University of Agriculture and Technology, P.O. Box 62000-00200, Nairobi, Kenya.

<sup>2</sup>Department of Biological Sciences, University of Embu, P.O. Box 6-60100, Embu Kenya.

## Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of coronavirus (CoV) disease 2019 (COVID-19). This study compared the genome, mutations, and infectivity/transmissibility of SARS-CoV-2 with selected betacoronaviruses (beta-CoVs). This study further examined the origin, risk factors, and outbreaks caused by beta-CoVs. We searched the following databases for relevant studies: PubMed, Google Scholar, and the World Health Organization COVID-19 database. A close relationship between SARS-CoV-2 and SARS bat-like CoV RaTG13 (98.9%) was found at the amino acid level, followed by pangolin CoVs. Non-synonymous mutations occur at high frequencies in the open reading frame (ORF) 1ab, spike (S) protein, and nucleocapsid. Mutations P323L and D614G in the RNA-dependent RNA polymerase (RdRp) and S protein, respectively, occur at a high frequency globally. Mutations at position 3037 in the nonstructural protein (Nsp) 3, 14408 (RdRp), and 23403 (S) confer transmissibility to SARS-CoV-2. SARS-CoV-2 has higher infectivity and transmissibility than SARS-CoV, which shares the same receptor. Although bats are confirmed reservoirs, intermediate hosts are currently unknown. Smoking, old age, diabetes, cardiovascular diseases, and hypertension have all been associated with COVID-19. Within six months of its outbreak, COVID-19 was reported in all countries worldwide, whereas SARS was reported in 28 countries and Middle East respiratory syndrome (MERS) in 5 countries. However, the fatality rate of MERS (65%) was higher than that of COVID-19 (4.9%) and SARS (6.6%). Identifying the SARS-CoV-2 intermediate hosts will help prevent future outbreaks. Attention should be given to the pangolin CoVs. Variations in the S gene may confer transmissibility and infectivity.

**Keywords:** Mutations, SARS-CoV-2, Pandemic, COVID-19, Coronaviruses, Respiratory Syndrome

\*Correspondence: mirriam.nzivo@jkuat.ac.ke

(Received: December 05, 2020; accepted: April 17 2021)

**Citation:** Nzivo MM, Budambula LM. Mutations and Epidemiology of SARS-CoV-2 compared to Selected Coronaviruses during the First Six Months of the COVID-19 Pandemic: A Review. *J Pure Appl Microbiol.* 2021;15(2):524-533. doi: 10.22207/JPAM.15.2.19

© The Author(s) 2021. **Open Access.** This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, sharing, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## INTRODUCTION

Betacoronaviruses (beta-CoVs) have been responsible for three outbreaks that are of public health significance over the past two decades. Coronavirus (CoV) disease 2019 (COVID-19) first appeared in December 2019, and the World Health Organization (WHO) announced the name of the new epidemic on February 11, 2020. The virus responsible was originally called the 2019 novel CoV. However, its name was changed by the International Committee on Taxonomy of Viruses to severe acute respiratory syndrome CoV2 (SARS-CoV-2)<sup>1,2</sup>. SARS-CoV-2 is a beta-CoV belonging to the subgenus *Sarbecovirus*. Other organisms in the genus of public health concern are SARS-CoV and Middle East respiratory syndrome (MERS)-CoV<sup>1</sup>. MERS-CoV caused the MERS outbreak in 2012–2015, resulting in 1,621 cases and 584 deaths<sup>3</sup>. SARS-CoV was responsible for the SARS outbreak, spreading across 37 countries in 2002–2003 resulting in 8,000 cases and 774 fatalities<sup>3,4</sup>. Other human CoVs (HCoVs) are NL63 and 229E of the *Alphacoronavirus* genus and OC43 and HKU1 of the genus *Gammacoronavirus*<sup>1,3,4</sup>.

## METHODS

Online databases, including PubMed, Google Scholar, and the WHO COVID-19 database, were searched for articles containing the phrase "SARS-CoV-2" and any of the following terms: "genome", "mutations", "epidemiology", "infectivity", "Spike protein" or "host". Additional search terms identified in the first round of screening were used to conduct subsequent searches in the databases. This included, "nCoV-2019" and "genome", "risk factors", "intermediate host" or "Spike protein". The articles obtained in each search were reviewed based on their titles and abstracts for relevance. The SARS-CoV-2 articles were published in January 2020. The WHO situation reports on SARS and MERS outbreaks and a systematic review of CoVs were also included.

### SARS-CoV-2 compared to selected coronaviruses

SARS-CoV-2 has approximately 29.9kb, non-segmented single-stranded positive RNA genome, enclosed in an envelope(E)<sup>5,6</sup>. At the beginning of the genome lies a short untranslated region (5' UTR) followed by an open reading frame 1 (ORF1), which is approximately 67% of the RNA. ORF 1 is composed of Pp1a and Pp1ab,

which encode 16 non-structural proteins involved in transcription and replication<sup>5,7-9</sup>. Nsp 12, also known as RNA-dependent RNA polymerase (RdRp), catalyzes RNA synthesis with Nsp 7 and Nsp 8 as co-factors<sup>7,10</sup>. At the 3' terminus, there are four structural proteins: the nucleocapsid (N), membrane (M), spike (S), and E<sup>8,11</sup>.

CoVs use S glycoproteins to gain entry into potential host cell<sup>8,11</sup>. The S1 region of the S protein allows binding where as the S2 region enhances fusion. Thus, the S protein determines host tropism<sup>8,12</sup>. Peptide insertion (PRRA) in the S1/S2 junction of the S protein (position 681–684) induces a furin cleavage motif (RRAR) and is unique to SARS-CoV-2<sup>13,14</sup>. M functions include nutrient transport, bud release, and the formation of envelope (E). In addition, together with the E protein, they participate in the assembly of viral particles. The N protein catalyzes RNA synthesis<sup>5,15</sup>. The SARS-CoV-2 genome also has eight accessory proteins: ORF 3a, 3b, 6, 7a, 7b, 8a, 9b, and 14, which are common to other CoVs<sup>5,16</sup>.

The genome of SARS-CoV-2 is more similar to the bat-SARS genome than the SARS-CoV and MERS-CoV genomes. For instance, SARS-CoV-2 is approximately 96.2% similar to the whole genome sequence of the bat CoV RaTG13 of *Rhinolophus affinis*, and 87%–89% similar to strains SL-CoVZC45 and SL-CoVZXC21 of *Rhinolophus sinicus*<sup>4,17</sup>. The SARS-CoV-2 genome is 85.5%–92.4% similar to that of Guangdong and Guangxipangolin SARS-related-CoVs<sup>18</sup>. SARS-CoV-2 has 50% and 79% whole genome sequence similarity to MERS-CoV and SARS-CoV, respectively<sup>18-20</sup>. SARS-CoV-2 is similar to SARS-CoV at the amino acid (aa) level but has distinct dissimilarities. The ORF 8 gene in SARS-CoV-2 is not segmented in 8a and 8b, similar to SARS-CoV<sup>22</sup>. ORF 3b is 22aa-long in SARS-CoV-2 where as it is 154aa long in SARS-CoV<sup>16</sup>. SARS-CoV-2 shares 40.2% similarity with mild respiratory HCoV-OC43 at the nucleotide level<sup>19</sup>. At the nucleotide level, the relatedness between SARS-CoV-2 and canine and bovine CoVs ranged from 40.8% to 41.5%. Comparisons between different genes showed that the S gene of SARS-CoV-2 has higher homology with that of bat CoVs (95.4%–97.7%) and pangolins (97.5%). The protein sequences of SARS-CoV-2 and SARS causative agent S genes derived from humans, civets, and bats have approximately 76%–78% similarity<sup>22-24</sup>.

The RdRp gene of SARS-CoV-2 is extremely similar (96%) to that of SARS-CoV.

We calculated the percent average of amino acid similarities of the whole genome, ORF 1ab, S, M, E, and N proteins of several CoVs to SARS-CoV-2. In case of a range, we used the highest value. Based on amino acid similarities of the entire genome and several regions (Table 1), a remarkably close relationship was observed between SARS-CoV-2 and bat-CoV RaTG13 (98.9%). This was followed by pangolin CoVs (96.2%), and SARS-CoV (85.2%). We also observed a distant relationship between SARS-CoV-2 and MERS-CoV genomes based on aa similarities (40.4%) of the whole genome, ORF 1ab, S, E, M, and N.

#### Mutations present on the SARS-CoV-2 genome

Similar to other RNA viruses, mutations in CoVs arise during every replication cycle<sup>4</sup>. Mutations in viruses can modulate virulence, transmission, receptor affinity, and host tropism among others<sup>6,25</sup>. Interestingly, in the order Nidovirales (to which Coronaviridae belongs) RNA polymerase has a proof reading mechanism. Nsp 14 in SARS-CoV has a proof reading function that increases the fidelity of RNA synthesis during exonuclease activity. It is expected to carry out the same function in SARS-CoV-2 correcting errors by RdRp<sup>26</sup>. Thousands of mutations, ranging from synonymous to non-synonymous, insertions, and deletions, have been reported in more than 50,000 genome sequences that were uploaded in the Global Initiative on Sharing all Influenza Data and the NCBI database. Approximately 80% of the reported mutations were found to be non-synonymous<sup>27</sup>.

We reviewed studies that compared 10 Chinese isolate sequences to more than 1,000

sequences that have been uploaded from around the world. Mutations have been found to occur at high frequencies in the ORF 1ab, S, and N genes, as well as ORF3a<sup>6,28,29</sup>. No mutations were detected in the Chinese isolates in December 2019<sup>4</sup>. The first non-synonymous mutation to be reported was in position 28151 in ORF 8, resulting in serine and leucine (S and L) clade. Position 28151 within ORF 8 is characterized by a cytosine or uracil that codes for a serine/leucine (S/L) variation at aa 84. Amino acid 84 is not conserved among the other CoVs. The S-Clade was the original virus but it became less prevalent in course of time (approximately 30%). The prevalence of the L-Clade was approximately 70% early in the outbreak but its frequency decreased in late January 2020. It was more aggressive than the S-clade<sup>5,18,30</sup>. The SARS-CoV-2 S-clade is characterized by a mutation at position 28151. The strains isolated in North America and Europe were of the L-variant descent<sup>26,29</sup>.

The other significant mutation was located at position 14408 of the RdRp gene. It is anticipated to have appeared first in Italy on February 20, 2020, resulting in an aa change P323L. At this time, the number of individuals contracting the virus increased substantially in Europe<sup>26,31</sup>. Viruses with the P14408L mutation have a higher mutation rate (median of three point mutations) than those without the mutation, with a median of one point mutation<sup>26</sup>. Clade G of SARS-CoV-2 is based on a mutation in the S gene at position 23403, resulting in D614G and became the dominant circulating variant as of June 2020. Clade V is based on a variation in position 26143 which resulted in aa change G251V in ORF 3a<sup>6,29</sup>. Subclades G.1 and G.2 are the result of the following mutations: G204R and R203K in N gene, P214L in ORF1b, and Q57H

**Table 1.** Amino acid sequence similarity in % of selected coronaviruses to SARS-CoV-2

	SARS-CoV	MERS-CoV	RaTG13	Pangolin CoVs	References
Whole genome	76	48	95.8-98.5	87.1-94.1	14,22,58-62
ORF 1ab	86.1	45.4	98.5	87.1-90.3	
Spike protein	75.9	28.9	91.5-97.4	88.3-97.5	
Envelope	94.7	35.3	100	97.4-100	
Membrane	89.1	38.7	98.1-100	93-98.6	
Nucleocapsid	90.2	45.9	99	96.7	
Average	85.3	40.4	98.9	96.2	

in ORF3a<sup>6</sup>. Mutations 2891 and 3036 (Nsp 3), 14408 (Nsp 12), 23403 (S protein), 28881G > A, 28882G > A, and 28883G > C (N) are characteristic of isolates from Europe<sup>26</sup>. Mutations 1059C>T, 18060 (Nsp 14), 17857, and 17746 (Nsp 13, helicase), and 25563G>T are the characteristics of North American isolates<sup>26,32</sup>. According to Pachetti et al. (2020)<sup>26</sup>, mutations at the position 241 in the leader sequence, 3037 (Nsp 3), 14408 (RdRp), and 23403 (S), confer transmissibility. Non-synonymous mutations 241 (leader sequence), L37F (Nsp 6), G251V, P323L, D614G (S gene), L84S (ORF 8), and synonymous mutations S75S (Nsp 4) and F105F (Nsp 3) occur in high-frequency globally<sup>6,29,33</sup>. Two deletions of three nucleotides and 24nt in the ORF 1ab and one deletion of 10nt at the 32 end of the genome were observed in

sequences of SARS-CoV-2 obtained from Aichi, Japan; Wisconsin, USA; and Victoria, Australia<sup>34</sup>. Table 2 summarizes the mutations that occur in different regions of the SARS-CoV-2 genome. The majority of mutations have been reported in ORF 1ab (n=59), S (n=32), and N (n=14). Figure 1 shows the logarithm (Log<sub>10</sub>) of mutations occurring in regions with more than five mutations.

**SARS-CoV-2 versus SARS-CoV infectivity**

The S glycoprotein of CoVs is crucial for infection and stability, and determines host tropism, as well as the transmission capacity of the virus<sup>4,8</sup>. The S protein attaches to the receptor cells, allowing an entry of the virus<sup>12,35</sup>. The S gene has a subunit, S1, which has a receptor-binding domain (RBD) that binds to angiotensin-converting enzyme 2 (ACE2), allowing the S2 domain to fuse

**Table 2.** Mutations occurring on the SARS-CoV-2 genome

Region	Genomic position	Non synonymous	Author
5'UTR		C241T	
ORF 1ab	266-21555	N=59 L36F, L37F, A88V, H116Q, P118L, A117T, A144V, P214L, T266I, H286Y, P287T, P308S, P 309S, P314L, P323L, G392D, S 428N, P504L, C541Y, T609I, T708I, I739V, P765S, H819Y, A876T, A1043V, A1176V, V1397I, L1599F, A1606T, I1607V, M2194T, L2235I, I2244T, G2251S, A2345 V, G2534V, D2579A, N2708S, N2894D, F2908 I, T3058I, F3071Y, S3099L, G3334S, L3606F, E3764 D, N3833K, L3691F, S4396L, W5308C, T5579I, I6074V, I 6075 T, P6083 L, F6309Y, E6565D, K6958R, D7018N	9,21,26,29,33,34,63,64
Spike Protein	21563-25384	N=32 Y28N, T29I, F32I, H48Y, H49, L54F, N74K, E96D, D111N, F157L, G181V, S221W, S 247 R, A348T, N354D, D 364Y, V366F, V367 F, R408, G476S, V483A, H519Q, A520S, D 614G, A930V, D936Y, A1078V, G1124V, P1143L, D1168H, E1207V, D1259H	9,26,28,34,56,64,65
ORF 3a	25393-26220	N=3 Q57H, G196V, I208T, G251V	33,64,66
Envelope	26245-26472	N=6 L37H, S50G, V52I, V58L, S68C, L73F	6
Matrix	26523-27191	N=3 T175M, D209H, I1208T	29,34,66
ORF 6	27202-27387		
ORF 7a	27394-27759		
ORF 7b	27756-27887		
ORF 8	27894-28259	N=3 V32L, V61L, L84S	(34, 40)
Nucleo-capsid	28274-29533	N=14 P46S, D103Y, T148I, S193I S194L, S197L, S202N, R203K, G204R, G238C, D249H, K343I, P344S, S404A	(22,30,37,38,40,41)
ORF 10	29558-29674	R9659H	67

and initiate an infection. The RBDs of SARS-CoV-2 and SARS-CoV have 89.2% sequence similarity and 70%–73.7% sequence identity with several amino acid variations in key residues. SARS-CoV-2, NL63, and SARS-CoV receptors use ACE2 as their main receptor<sup>10,49</sup>, whereas MERS-CoV attaches to dipeptidyl peptidase 4.

More RBD residues are involved in binding to the receptor, and they have a longer capping loop in SARS-CoV-2 than in SARS-CoV. These structural characteristics are associated with an increased affinity for binding<sup>8</sup>. The SARS-CoV-2 S protein binds to ACE2 with significantly higher affinity than that of SARS-CoV. Therefore, it is more efficient at infecting human cells and hence, rapidly spreads in the human population<sup>8,36,38,39</sup>. The binding free energy between RBD-ACE2 in SARS-CoV-2 is lower than that of SARS-CoV and explains why it is more infectious than the latter. This indicates that it has increased stability and can withstand high temperatures<sup>8,35</sup>. The SARS-CoV-2 RBD domain has higher solubility than the SARS-CoV domain, enhancing binding to the ACE2 receptor and contributing to its high infectivity<sup>8</sup>. Besides SARS-CoV-2, the rate of protein synthesis is higher than that of HCoV-229E due to the overall ratios of human slow codons/di-codons present, thereby increasing the transmissibility rate. Fast-replicating viruses have a high chance of survival in the environment and have successful infection<sup>31</sup>.

The mutations that occur in SARS-CoV-2 residues, responsible for cross-species transmission, could be more favorable than those in SARS-CoV<sup>31,36</sup>. Peptide insertion at the S1/S2 boundary has been attributed to SARS-CoV-2 infectivity<sup>14,16</sup>. A previous study<sup>40</sup> reported that the S protein of SARS-CoV-2 weakly binds to its receptor, unlike SARS-CoV, justifying its severity only in immunocompromised individuals. However, this does not explain the extensive spread of SARS-CoV-2 infection within a short duration. This evidence is also supported by the trends of SARS-CoV-2 transmission, with more than 10 million people being infected within a period of six months since its outbreak.

#### Intermediate host

COVID-19 was first reported in nine patients with pneumonia in China in December 2019. Of the nine patients, eight had visited the Huanan wet market in Wuhan City and the other patient lived close to the market before the outbreak<sup>4,41</sup>. Huanan wet markets sell different wild animal species, including marmots, pangolins, snakes, leopard cats, bamboo rats, badgers, and hedgehogs. All of these species are susceptible to CoVs. SARS-CoV-2 is suspected to have originated from bats as it shares a 87%–96% whole genome sequence identity with bat CoVs, such as RaTG13, SL-CoVZC45, and SL-CoVZXC21<sup>11,20,22</sup>. Based on the analysis of complete genomic sequences, bats are reservoirs of approximately 30 CoVs. However,

**Table 3.** The first six months of COVID-19, MERS and SARS disease outbreaks

Organism	SARS-COV-2	SARS-COV	MERS-CoV	References
Disease	Covid-19	SARS	MERS	2,3
First case presented in	China	China	Saudi Arabia	2,3
Overall Case fatality in %	4.9	6.6	65	
Incubation Period in days	2-14	2-11	2-13	3,46
Median to time of death in days	14	23	14	3,46
Cases	10 185 374	5663	17	68-70
Deaths	503862	372	11	68-70
Transmission mode	Droplets, close contact	Droplets, close contact	Droplets, close contact	3,39
Human-human transmission	Efficient	Efficient	Limited	71
Pandemic potential	Yes	Yes	No	71
Distribution in countries	213	28	5	68-70
Reservoir	Bats	<i>Rhinolophus sinicus</i>	Bats and alpacas	3,39
Intermediate host	Pangolins (to be confirmed)	Palm civets	Dromedary camel	3,39,68-70

during the out break period, no bats were being sold as most species had hibernated. Therefore, another organism was likely to be an intermediate host<sup>4</sup>.

Several animals, including pangolins, snakes, and turtles, are suspected to be SARS-CoV-2 intermediate hosts<sup>22,42-44</sup>. Although illegally sold, pangolins are a delicacy in China, and their scales are used in Chinese traditional medicine<sup>45</sup>. At the nucleotide level, pangolin CoVs and SARS-CoV-2 share 85.5%–99% sequence similarity<sup>22,44,45</sup>. Pangolin-CoVs are also closely related at the whole genome level to bat CoV RaTG13 (90.55%), SARSr-CoV ZXC21(85.65%), and bat SARSr-CoV ZC45 (85.01%)<sup>22</sup>. SARS-CoV-2 has an insertion, PRAR at the S1 and S2 sites, which induces a furin cleavage motif that is absent in pangolin CoVs. Therefore, although pangolin-CoVs are genetically closely related to SARS-COV-2 and bat CoVs, pangolins may not be intermediate hosts<sup>12,16</sup>.

Ji et al.<sup>42</sup> reported that snakes of the species *Bungarus multicinctus* and *Naja atra* are potential intermediate hosts. However, this was refuted by another study that showed that the relative synonymous codon usage distance was smaller between SARS-CoV-2, and a frog species than these snake species<sup>23</sup>. However, they did not suggest frogs as the intermediate host but further argued that the intermediate host would most likely be a mammal. Turtles were also suspected to be a possible intermediate because of the observed interaction between RBD key aa and their ACE2. In addition, they were reported to be

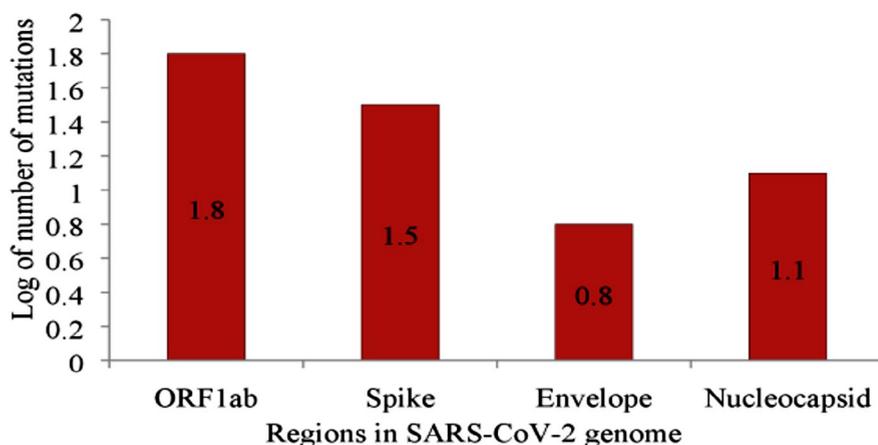
more common in the market<sup>12</sup>. However, more research is required to determine this possibility.

#### Risk factors

Several sociodemographic characteristics were associated with COVID-19. SARS-CoV-2 infects all age groups with a median age of 45–56 years. Patients aged above 45 years respond poorly to treatment; hence, they have increased fatalities<sup>5,46</sup>. COVID-19 is also more prevalent in men than in women, with a mortality rate 2.4 times higher among men than among women<sup>47-49</sup>. The high incidence observed among men could be a result of higher prevalence of smoking and movement, including international travel, compared to women in many cultures.

Behavioral characteristics, such as human–animal contact, poor hygiene and sanitation, and wet markets increase the risk of zoonotic diseases<sup>50</sup>. Smoking increases the susceptibility and death rate of SARS-CoV-2 infection<sup>5,51</sup>. ACE2 expression levels in long-term smokers are up regulated<sup>51</sup>. Normally, ACE2 is upregulated to protect the host against acute lung injury. The levels of ACE2 do not vary by sex or age. Chronic smoking triggers ACE2 secretory cell expansion, which could explain the vulnerability as SARS-CoV-2 uses the ACE2 receptor<sup>51,52</sup>.

A temperature range of 13–24°C, precipitation of less than 30mm/month, and humidity of 50%–80% favors SARS-CoV-2 infection. Lower rainfall reduces relative humidity and provides favorable conditions for the spread of pathogens causing respiratory infections<sup>53</sup>. Several



**Fig. 1.** Log<sub>10</sub> non-synonymous mutations occurring in genome of SARS-CoV-2

studies support this hypothesis,<sup>54</sup> reporting that an increase in temperature and absolute humidity is associated with decreased mortality after observing weather conditions in relation to the death rate for one month. The RBD-ACE2 complex results in a high entropy penalty owing to the flexibility of the RBD near the binding site. This means that it is temperature-sensitive regarding infections in humans. Therefore, an increase in temperature decreases the rate of infection, and can therefore be controlled<sup>8</sup>. The increased fatality rate of SARS-CoV-2 is correlated with the levels of particulate matter (PM) pollution (PM<sub>2.5</sub>, PM<sub>10</sub>). PM affects the outcome of respiratory diseases<sup>55</sup>.

Underlying conditions that worsen COVID-19 outcomes include diabetes mellitus, hypertension, and cardiovascular diseases<sup>35,39,46</sup>. Among patients with diabetes and cardiovascular diseases, ACE2 plays a protective role. Unfortunately, SARS-CoV-2 downregulates ACE2 protein, resulting in severe clinical outcomes in these patients. Both, patients with respiratory diseases, as well as healthy individuals, are susceptible to SARS-CoV-2 infection<sup>51</sup>. Both healthy individuals and those with chronic respiratory diseases showed similar ACE2 expression levels<sup>51</sup>. The recurrence of positive SARS-CoV-2 RNA predictors includes high levels of IL-6, elevated lymphocyte counts, and lung consolidation features upon hospital admission<sup>11,39,51</sup>. The transmission of SARS-CoV-2 in countries with malaria incidence is limited<sup>56</sup>. This could be a result of the wide use of antimalarial drugs, especially chloroquine. Chloroquine is a broad-spectrum antiviral agent that increases endosomal pH, which is essential for enhancing the fusion of the virus and host cells. It inhibits the uncoating and glycosylation of many viruses. Chloroquine in combination with other antiviral agents was reported to be effective in controlling COVID-19<sup>56,57</sup>.

### The outbreaks

COVID-19 was first reported on December 31, 2019, SARS on November 1, 2002, and MERS on September 20, 2012 but was traced back to April 2012. Therefore, this study focused on the first six months of the outbreak: COVID-19 on June 30, 2020, SARS upto April 30, 2003, and MERS upto February 20, 2013. By June 30, 2020, approximately 10 million people had contracted COVID-19 resulting in approximately 500,000

fatalities. This is more than 1,000 times as many people that contracted SARS and MERS and died during the six months after their outbreaks (Table 3). Of the three outbreaks, COVID-19 is the most extensively spread and was reported in over 200 countries within the first six months of its emergence; in contrast, MERS and SARS were reported in 5 and 28 countries, respectively. However, the COVID-19 mortality rate within the first six months of the outbreak was 4.9%, indicating that it is less aggressive than SARS (6.6%) and MERS (65%). The etiological agents for these outbreaks have been linked to bats with different intermediate hosts.

### CONCLUSION

We reviewed studies that reported the genomes, mutations and epidemiology of the pathogen responsible for COVID-19 and the number of selected CoVs. The SARS-CoV-2 genome is highly similar to the strain RaTG13 of bat origin. This was followed by pangolins based on amino acid similarities of the whole genome, S, M, E, and N. Mutations occurred at a high frequency in important proteins, namely the S, N, and replicase. ORFs 6, 7, and 10 were highly conserved. Insights into the sequence variations, especially important genes, such as the S glycoprotein, are important for understanding the biology of SARS-CoV-2 infection. This knowledge could be useful for antiviral treatment and vaccine development. Although SARS-CoV-2 and SARS-CoV use the same receptor in humans, the former has high binding affinity to the receptor. This results in high infectivity and transmissibility rates. Despite evidence of a close relationship between pangolins and SARS-CoV-2 in most studies, there remains an uncertainty regarding their identification as intermediate hosts. The risk factors for COVID-19 have been defined and are similar to those of other respiratory diseases; smoking and air pollution have been associated with disease severity. However, it has not been associated with cold weather, and this could partly explain the mortality variations in different regions. SARS-CoV-2 is highly transmissible and infectious, with a low mortality rate. We recommend further research to identify SARS-CoV-2 intermediate hosts to avoid future spillover, and to establish the significance of different mutations.

## ACKNOWLEDGMENTS

None.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest

## AUTHOR CONTRIBUTION

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

None.

## DATA AVAILABILITY

All data generated or analyzed during this study are included in the manuscript.

## ETHICS STATEMENT

Not applicable.

## REFERENCES

- Gorbalenya AE, Baker SC, Baric RS, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 2020;5(4):536-544. doi:10.1038/s41564-020-0695-z
- World Health Organization (WHO). Novel coronavirus - China 11-fev 2020. WHO Bull. 2020;(February). [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1\\_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2).
- Su S, Wong G, Shi W, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* 2016;24(6):490-502. doi:10.1016/j.tim.2016.03.003
- Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet.* 2020;395(10224):565-574. doi:10.1016/S0140-6736(20)30251-8
- Guo YR, Cao QD, Hong ZS, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- A n update on the status. *Mil Med Res.* 2020;7(1):1-10. doi:10.1186/s40779-020-00240-0
- Kim J-S, Jang J-H, Kim J-M, Chung Y-S, Yoo C-K, Han M-G. Genome-Wide Identification and Characterization of Point Mutations in the SARS-CoV-2 Genome. *Osong Public Heal Res Perspect.* 2020;11(3):101-111. doi:10.24171/j.phrp.2020.11.3.05
- Cui H, Gao Z, Liu M, et al. Structural genomics and interactomics of 2019 Wuhan novel coronavirus, 2019-nCoV, indicate evolutionary conserved functional regions of viral proteins. *bioRxiv.* 2020:2020.02.10.942136. doi:10.1101/2020.02.10.942136
- He J, Tao H, Yan Y, Huang SY, Xiao Y. Molecular mechanism of evolution and human infection with SARS-CoV-2. *Viruses.* 2020;12(4):1-19. doi:10.3390/v12040428
- Morais IJ, Polveiro RC, Souza GM, et al. The global population of SARS-CoV-2 is composed of six major subtypes. *Sci Rep.* 2020;10(1):2-10. doi:10.1038/s41598-020-74050-8
- Gao Y, Yan L, Huang Y, et al. Structure of RNA-dependent RNA polymerase from 2019-nCoV, a major antiviral drug target. *bioRxiv.* 2020:2020.03.16.993386. doi:10.1101/2020.03.16.993386
- Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol.* 2020;92(4):418-423. doi:10.1002/jmv.25681
- Liu Z, Xiao X, Wei X, et al. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J Med Virol.* 2020;:-7. doi:10.1002/jmv.25726
- Li X, Wang W, Zhao X, et al. Transmission dynamics and evolutionary history of 2019-nCoV. *J Med Virol.* 2020;92(5):501-511. doi:10.1002/jmv.25701
- Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-. *Cell.* 2020;180:1-12.
- Tiwari M, Mishra D. Investigating the genomic landscape of novel coronavirus (2019-nCoV) to identify non-synonymous mutations for use in diagnosis and drug design Manish. 2020:1-5. doi:10.3969/j.issn.1001-1242.2020.01.020
- Wu A, Peng Y, Huang B, et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe.* 2020;27(3):325-328. doi:10.1016/j.chom.2020.02.001
- MacKenzie JS, Smith DW. COVID-19: A novel zoonotic disease caused by a coronavirus from China: What we know and what we don't. *Microbiol Aust.* 2020;41(1):45-50. doi:10.1071/MA20013
- Tang X, Wu C, Li X, et al. On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev.* 2020;7(6):1012-1023. doi:10.1093/nsr/nwaa036
- Malik YS, Sircar S, Bhat S, et al. Emerging novel coronavirus (2019-nCoV)—current scenario, evolutionary perspective based on genome analysis and recent developments. *Vet Q.* 2020;40(1):68-76. doi:10.1080/01652176.2020.1727993
- Fahmi M, Kubota Y, Ito M. Nonstructural proteins NS7b and NS8 are likely to be phylogenetically associated with evolution of 2019-nCoV. *Infect Genet Evol.* 2020;81:104272. doi:10.1016/j.meegid.2020.104272
- Benvenuto D, Demir AB, Giovanetti M, et al. Evidence for mutations in SARS-CoV-2 Italian isolates potentially affecting virus transmission. *J Med Virol.* 2020:0-3. doi:10.1002/jmv.26104
- Zhang T, Wu Q, Zhang Z. Pangolin homology associated with 2019-nCoV. *bioRxiv.* 2020:2020.02.19.950253. doi:10.1101/2020.02.19.950253
- Zhang C, Zheng W, Huang X, Bell EW, Zhou X, Zhang Y. Protein Structure and Sequence Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique Similarity between Its Spike Protein

- Insertions and HIV-1. *J Proteome Res.* 2020;19(4):1351-1360. doi:10.1021/acs.jproteome.0c00129
24. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J Virol.* 2020;94(e00127-20):1-9. doi:10.1128/jvi.00127-20
25. Dilucca M, Pavlopoulou A. Analysis of codon usage and evolutionary rates of the 2019-nCoV genes. *bioRxiv.* 2020;7:2020.03.25.006569. doi:10.1101/2020.03.25.006569
26. Pachetti M, Marini B, Benedetti F, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med.* 2020;18(1):1-9. doi:10.1186/s12967-020-02344-6
27. van Dorp L, Acman M, Richard D, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol.* 2020;83(May):104351. doi:10.1016/j.meegid.2020.104351
28. Lokman SM, Rasheduzzaman M, Salaudin A, et al. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: A computational biology approach. *Infect Genet Evol.* 2020;84. doi:10.1016/j.meegid.2020.104389
29. Yin C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics.* 2020;(January). doi:10.1016/j.ygeno.2020.04.016
30. Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. *J Med Virol.* 2020;92(5):522-528. doi:10.1002/jmv.25700
31. Yang CW, Chen MF. Composition of human-specific slow codons and slow di-codons in SARS-CoV and 2019-nCoV are lower than other coronaviruses suggesting a faster protein synthesis rate of SARS-CoV and 2019-nCoV. *J Microbiol Immunol Infect.* 2020;(xxx):0-5. doi:10.1016/j.jmii.2020.03.002
32. GitHub - nextstrain/ncov: Nextstrain build for novel coronavirus SARS-CoV-2. <https://github.com/nextstrain/ncov>.
33. Ugurel OM, Ata, Oguz, Turgut-Balik D. An updated analysis of variations in SARS-CoV-2 genome. *Turkish J Biol.* 2020;44:157-167. doi:10.3906/biy-2005-111
34. Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol.* 2020;81(January). doi:10.1016/j.meegid.2020.104260
35. He F, Deng Y, Li W. Coronavirus disease 2019: What we know? *J Med Virol.* 2020;2019:0-2. doi:10.1002/jmv.25766
36. Ortega JT, Serrano ML, Pujol FH, Rangel HR. Role of Changes in SARS-COV-2 Spike Protein in the Interaction with the Human Ace2 Receptor: An In Silico Analysis. *Excli J.* 2020;19(19):410-417. doi:10.17179/excli2020-1167
37. Cao Y, Li L, Feng Z, et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discov.* 2020;6(1):4-7. doi:10.1038/s41421-020-0147-1
38. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020. doi:10.1038/s41586-020-2180-5
39. Xu J, Zhao S, Teng T, et al. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses.* 2020;12(2). doi:10.3390/v12020244
40. Dong N, Yang X, Ye L, Chen K, Chan EW-C, Chen S. Genomic and protein structure modelling analysis depicts the origin and pathogenicity of 2019-nCoV, a new coronavirus which caused a pneumonia outbreak in Wuhan, China. *F1000Research.* 2020;9:121. doi:10.12688/f1000research.22357.2
41. Zhou P, Yang X, Wang X, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020;579(March). doi:10.1038/s41586-020-2012-7
42. Ji W, Wang W, Zhao X, Zai J, Li X. Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. *J Med Virol.* 2020;92(4):20200122. doi:10.1002/jmv.25682
43. Li X, Zai J, Zhao Q, et al. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. *J Med Virol.* 2020:1-10. doi:10.1002/jmv.25731
44. Liu Ping, Jiang Jing-Zhe, Wan Xiu-Feng, Hua Yan, Wang Xiaohu, Hou Fanghui, Chen Jing, Zou Jiejian CJ. Are pangolins the intermediate host of the 2019 novel coronavirus (2019-nCoV) ? *bioRxiv Prepr.* 2020:1-9. doi:10.1016/j.solener.2019.02.027
45. Lam TT-Y, Shum MH-H, Zhu H-C, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature.* 2020:1-6. doi:10.1038/s41586-020-2169-0
46. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun.* 2020;(February):18-21. doi:10.1016/j.jaut.2020.102433
47. Coronavirus Age, Sex, Demographics (COVID-19) - WorldometerNo Title. <https://www.worldometers.info/coronavirus/coronavirus-age-sex-demographics/>.
48. Jin JM, Bai P, He W, et al. Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Front Public Heal.* 2020;8(April):1-6. doi:10.3389/fpubh.2020.00152
49. World Health Organization (WHO). Gender and COVID-19. 2020;(May):6-9.
50. Li HY, Zhu GJ, Zhang YZ, et al. A qualitative study of zoonotic risk factors among rural communities in southern China. *Int Health.* 2019;12(2):77-85. doi:10.1093/inthealth/ihaa001
51. He X, Zhang L, Ran Q, et al. Integrative Bioinformatics Analysis Provides Insight into the Molecular Mechanisms of 2019-nCoV. 2020. doi:10.1159/000488830
52. Smith JC, Sausville EL, Girish V, et al. Cigarette Smoke Exposure and Inflammatory Signaling Increase the Expression of the SARS-CoV-2 Receptor ACE2 in the Respiratory Tract. *Dev Cell.* 2020;53(5):514-529.e3. doi:10.1016/j.devcel.2020.05.012
53. Bu J, Peng D-D, Xiao H, et al. Analysis of meteorological conditions and prediction of epidemic trend of 2019-nCoV infection in 2020. *medRxiv.* 2020;(May):2020.02.13.20022715. doi:10.1101/2020.02.13.20022715
54. Ma Y, Zhao Y, Liu J, et al. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Sci Total Environ.* 2020;724:138226.

- doi:10.1016/j.scitotenv.2020.138226
55. Yao Y, Pan J, Liu Z, et al. Temporal Association Between Particulate Matter Pollution and Case Fatality Rate of COVID-19 in Wuhan, China. *medRxiv*. 2020:2020.04.09.20049924. doi:10.1101/2020.04.09.20049924
  56. Bajaj A, Purohit HJ. Understanding SARS-CoV-2: Genetic Diversity, Transmission and Cure in Human. *Indian J Microbiol*. 2020. doi:10.1007/s12088-020-00869-4
  57. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents*. 2020;55(3):105924. doi:10.1016/j.ijantimicag.2020.105924
  58. Benvenuto, D., Giovanetti, M., Ciccozzi, A., Spoto, S., Angeletti, S., & Ciccozzi, M. (2020). The 2019-new coronavirus epidemic: Evidence for virus evolution. *Journal of Medical Virology*, 92(4), 455–459. <https://doi.org/10.1002/jmv.25688>
  59. Chan, J. F. W., Kok, K. H., Zhu, Z., Chu, H., To, K. K. W., Yuan, S., & Yuen, K. Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes and Infections*, 9(1), 221–236. <https://doi.org/10.1080/22221751.2020.1719902>
  60. Xiao K, Zhai J, Feng Y, et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature*. 2020;583(7815):286-289. doi:10.1038/s41586-020-2313-x
  61. Xu Y. Genetic diversity and potential recombination between ferret coronaviruses from European and American lineages. *J Infect*. 2020;80(3):350-371. doi:10.1016/j.jinf.2020.01.016
  62. Zhou Y, Zhang S, Chen J, Wan C, Zhao W, Zhang B. Analysis of variation and evolution of SARS-CoV-2 genome. *J South Med Univ*. 2020;40(2):152-158. doi:10.12122/j.issn.1673-4254.2020.02.02
  63. Poterico JA, Mestanza O. Genetic variants and source of introduction of SARS-CoV-2 in South America. *J Med Virol*. 2020:0-1. doi:10.1002/jmv.26001
  64. Chaw SM, Tai JH, Chen SL, et al. The origin and underlying driving forces of the SARS-CoV-2 outbreak. *J Biomed Sci*. 2020;27(1):73. doi:10.1186/s12929-020-00665-8
  65. Maitra A, Sarkar MC, Raheja H, et al. Mutations in SARS-CoV-2 viral RNA identified in Eastern India: Possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. *J Biosci*. 2020;45(1):1-18. doi:10.1007/s12038-020-00046-1
  66. Wang C, Liu Z, Chen Z, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol*. 2020;92(6):667-674. doi:10.1002/jmv.25762
  67. Benvenuto D, Angeletti S, Giovanetti M, et al. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. *J Infect*. 2020;81:e24-e27. doi:10.1016/j.jinf.2020.03.058
  68. WHO | Update 43 - Situation in China: more cases than the rest of the world combined. [https://www.who.int/csr/sars/archive/2003\\_04\\_30/en/](https://www.who.int/csr/sars/archive/2003_04_30/en/).
  69. World Health Organization (WHO). Coronavirus disease, Situation Report -162. 2020;(June).
  70. Zumla AI, Memish ZA. Middle East respiratory syndrome coronavirus: epidemic potential or a storm in a teacup? *Alimuddin*. *Eur Respir J*. 2014;43:1243-1248. doi:10.1183/09031936.00227213
  71. Lee Ping-Ing, Hsueh P-R. Emerging threats from zoonotic coronaviruses—from SARS and MERS to 2019-nCoV. *J Microbiol Immunol Infect*. 2020;(January):19-21.