GENETICS OF SARS-COV-2 AND GENETIC DETERMINANTS OF PATHOGENICITY

Dhurata BOZO
University of Sports, Tirana, Albania
Ilia MIKEREZI
Academy of Science, Tirana, Albania

ABSTRACT

The Severe Acute Respiratory Syndrome-2 virus (SARS-CoV-2) is a SARS-like beta-coronavirus of zoonotic origin, first identified in December 2019 in Wuhan, China. The first complete sequence of the genome was deposited at NCBI Genbank on January 5, 2020. Human-to-human transmission was confirmed on January 14, 2020, at which time SARS-CoV-2 had spread in many countries around the world. RNA viruses explicit high mutation rates, originating different genetic variants and assemblies of the viral genome which can affect the virus' infectivity, immune response, symptoms and clinical manifestations in patients. Thousands of global SARS-CoV-2 assemblies available in concentrated genetic databases are already known globally. Coronaviruses, including SARS-CoV-2, are single-positive-stranded, + ssRNA virus). Similarly to other coronaviruses, the SARS-CoV-2 genome consists of approximately 30,000 nucleotides organized into specific genes encoding structural proteins and non-structural proteins (Nsps). Structural proteins include the spike S), envelope (E), membrane (M), and nucleocapsid protein (N). Surface glycoprotein S is involved in binding to the host cell and plays an important role in rapid human-to-human transmission. Non-structural proteins enable the viral replication and transcription inside the host cell. The infection mechanism of SARS - CoV - 2 into human cells consists on the recognition and binding to ACE2 receptors of the human cell membrane, and entrance of the virus into the cell, enabled by intracellular proteases TMPRSS2. Thus, all human cells, tissues or organs that express, produce and release high levels of ACE2 are potential targets of SARS - CoV 2 infection. The target cells of SARS-CoV-2, express high levels of ACE-2 gene, determining the pathogenic effects in a variety of tissues and organs, including the most severe acute respiratory syndrome. Comparative genetic analyzes have revealed that SARS-CoV-2 is similar in 94.6% of the amino acid sequence and 80% of the nucleotide sequence to the previous SARS-CoV, a similarity to consider it a variant of SARS, hence its designation as SARS-CoV-2. Recent studies regarding mutations and rapid adaptation of the coronaviruses to the host show that genetic characteristics common to SARS-CoV-2 and the previous coronaviruses SARS-CoV and MERS-CoV affect their

virulence, high pathogenicity, the transmission capacity, the host susceptibility, the mutational ability and be zoonotically transmitted to humans. Recent studies show that the degree of severity of COVID-19 disease depends on human genetics, and specifically on the genetic variants/alleles of candidate genes such as ACE2, ADAM17 and TMPRSS2. Population genetic studies show clear correlations between the ACE2 expression level and the severity of COVID-19. On the other hand, ACE2 release from human cell is affected by other genes like ADAM17 and TMPRSS2. Genetic population studies show that their spread in different populations and ethnicities, may explain why severity and mortality rate is higher in males, why they affect individuals suffering from other concomitant diseases and why it does not show severe clinical symptoms in children. These results lead to hypotheses that the impact and risk of SARS-Cov-2, the severity of the disease and the mortality rate have a multifactorial genetic basis and that large-scale individual genetic profiling for these genes are required to identify the most genetically-vulnerable and at-risk populations.
Keywords: SARS-CoV-2, COVID-19, S protein, ACE-2 receptor, SARS etiology, ADAM-17 gene, TMPRSS-2 gene

1. INTRODUCTION

SARS-CoV-2 is a SARS-like coronavirus of potential zoonotic origin first identified in December 2019 in Wuhan, China's Hubei province. The etiological agent was characterized as a SARS-like betacoronavirus, which full genome was deposited at the NCBI Genebank on January 5, 2020. Thousands of genomes have been sequenced since that date. Trasmission from animals to humans was confirmed on 14 January 2020, at which time SARS-CoV-2 had already spread rapidly to many countries around the world. Further widespread global transmission led the WHO to declare COVID-19 as a pandemic on March 11, 2020. Due to its worldwide distribution, SARS-CoV-2 exhibits a wide genetic diversity, with considerable distribution of its forms and variants in different racial/ethnic nationalities. Genetic studies in populations with geographically diverse backgrounds have reported significant genetic variation in protein coding regions, with variations in allele/genotype frequencies. This data enriches our knowledge of SARS-CoV-2 and previously known coronaviruses to understand how the virus is adapting to its new host, in order to design and produce effective drugs and vaccines especially in pandemic conditions (JHCHS 2020; Van Dorpa *et al*., 2020).

Analysis of genetic sequence data of viral pathogens are an important tool for the epidemiological studies of of infectious diseases and the molecular orientation in the production of drugs and vaccines. Data on genetic structure and sequence shed light on key epidemiological parameters such as the timing of outbreak of epidemic doubling, transmission pathogen detection and identification of potential sources and / or animal reservoirs. On the other hand, genomic data are effectively used to identify pathogenic genes that

interact with the host and characterize the variable genetic zones, in order to avoid rapid mutations that reduce the effectiveness of drugs and vaccines (Fani *et al*., 2020).

This publication is a summary review on issues and scientific aspects related to the current state of knowledge in the fields related to SARS-CoV-2/COVID-19, pathogenicity, medical and pathological aspects and public health approaches, as reported from reliable and reputable sources and publications in scientific and informative journals until December 2020.

SARS-CoV-2: genomics

Coronaviruses, including SARS-CoV-2, are single-stranded, positive-stranded RNA viruses (+ ssRNA virus). Coronaviruses contain among the largest genomes among RNA viruses: 25-32 Kb. The SARS-CoV-2 genome consists of approximately 30,000 nucleotides organized into specific genes encoding structural proteins and non-structural proteins (Nsps). The typical CoV genome includes a 5'-cap region, 5'-untranslated region (UTR) , open reading frames, a 3'-UTR and 3'-poly (A) end edge. The first two-thirds of the genome encodes unstructured proteins from 2 open reading frames; enzymatic replicase complex. The last third of the genome encodes 4 well-conserved structural proteins in CoV. Structural proteins include spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins. Surface glycoprotein S is involved in interaction with the host cell angiotensin converting enzyme (ACE2) receptor 2 and plays an important role in rapid human-to-human transmission. S protein is responsible for binding to host cell receptors and viral entry into host cells. Proteins M, E and N are part of the viral particle nucleocapsid (JHCHS, 2020, Wang et al., 2020).

Non-structural proteins, which are generated as cleavage products of the open reading frame 1ab (ORF1ab) viral polyproteins, are assembled and enable viral replication and transcription. RNA-dependent RNA polymerase, also known as Nsp12, is the major component that regulates viral RNA synthesis with the help of Nsp7 and Nsp8. The other five accessory proteins encoded by the genes ORF3a, ORF6, ORF7a ORF8 and ORF10, assist in the final assembly of viral particles and their exit from host cells (Van Dorpa *et al*., 2020; Bahrami *et al*., 2020).

RNA viruses exhibit high mutation rates, thus creating different genetic versions of the viral genome in almost every replication cycle within the host cells. This process originates viral populations with a high diversity of genomes, known as *quasispecies*. During each viral replication cycle, the differences increase and accumulate, creating significant genetic differences between the original and the succeeding viral genome. Depending on the type and location of the mutations, genetic changes of this type can affect the infectivity, the trasmission capacity of the virus, the human immune response,

symptoms and clinical manifestations in patients (Zeng et al., 2020; Wang et al., 2020).

There are thousands of SARS-CoV-2 sequences available in the global genetic database. The extraordinary availability of genomic data reported during the COVID-19 pandemic has been made possible by the intensive and extraordinary efforts and studies of hundreds of researchers who are constantly discovering new sequence and deposit SARS-CoV-2 'assemblies' data globally (Wang *et al*., 2020; Hu *et al*., 2020).

Comparative sequence analysis of samples from infected individuals revealed that SARS-CoV-2 and SARS-CoV are very similar; the 2 viruses are 94.6% similar in amino acid sequence, while the nucleotide sequence similarity is 80% throughout the genome. For this reason in February 2020, the same Coronavirus Study Group classifies it as belonging to the same spacies and officially named the new coronavirus as SARS-CoV-2. While the clinical presentation, epidemiological patterns, and host range of SARS-CoV-2 may differ from the original SARS-CoV, it is the genetic similarity between the 2 viruses which led the group to classify them as same species. According to World Health Organisation the designation "SARS-CoV-2" is associated with the name of the virus, while "COVID-19" refers to the name of the disease caused by infection by SARS-CoV-2 (Hu *et al*., 2020; Rabaan *et al*., 2020).

SARS-Cov-2: infection mechanism

CoV binding to the host cell surface receptors and membrane fusion is mediated by the Spike (S) protein, composed of two subunits, S1 and S2. In the case of SARS - CoV - 2, the cleavage and activation of S proteins are controlled by the TMPRSS2 intracellular protease. This promotes the initial entry of the virus inside the host cell. Besides the changes in the aminoacid sequences of the SARS-CoV-2 and SARS-CoV S proteins, both use the same cellular receptor for entering into the cell: the angiotensin-converting enzyme-2 (ACE2) receptor. SARS protein CoV- 2 has high affinity for the ACE2; the RBD-S complex with ACE2 further promotes the binding of S2. S2 is higly expressed in SARS-CoV-2, compared to SARS, which may explain the higher infectivity of SARS-CoV-2. In particular, two specific repetitions of S2 subunit play the key role for fusion between the virus and the human cell cell membrane (Lotfi *et al*., 2020; Sironi *et al*., 2020).

All cells, tissues and organs with high ACE2 expression are potential susceptible targets of SARS - CoV 2 infection. Liver damage, intestinal inflammation, renal and testicular insufficiency and pancreatitis in patients with COVID-19, correlate closely with high expression of ACE2 in cholangiocytes, in the gastrointestinal tract (both small intestine and duodenum), urinary organs (kidneys), testicles and pancreas. ACE2 represents

a high level of expression in the heart, which can contribute to acute myocardial damage and chronic damage to the cardiovascular system. Although ACE2 is primarily expressed in human type II (AT2) alveolar cells, the ratio of type I and type I alveolar cell surface area II in alveoli is about 9:1. Type II alveolar cells expressing ACE2 occupy only 1.4% of all AT2 cells, which hypothesizes existence and inclusion of other receptor-assisted or unknown molecular interactions in regulating of SARS-CoV-2 infection and explaining its high infectivity. Other potential receptors, such as DC - SIGN (CD209 genotypes), L - SIGN (CLEC4M) and CD147 may also allow the virus to either capture and anchor in host cells or bind to the S protein. However, the role of these receptors and other influencing factors are still under investigation (Lotfi *et al*., 2020; Nakagawa and Miyazawa 2020).

SARS-CoV-2; human pathogenicity

Coronaviruses (CoV) are a large family of viruses that are spread from animal hosts to humans, causing life-threatening respiratory illnesses such as Middle East Respiratory Syndrome (MERS) and severe acute respiratory syndrome (SARS). The SARS-CoV-2 virus is genetically closely linked to the SARS-CoV coronavirus, the first pandemic threat of a new and deadly coronavirus that appeared in late 2002 and caused an outbreak of severe acute respiratory syndrome (SARS). SARS-CoV was extremely deadly but was extinguished after major public health mitigation measures. SARS-CoV-2 has spread rapidly worldwide compared to 2002 SARS-CoV and 2012 Middle East Respiratory Syndrome (MERS-CoV) coronavirus, although the estimated fatality rate in confirmed cases is 6.6% in SARS-CoV-2 (as of August 2020), which is lower than those of SARS-CoV and MERS-CoV, (9.6% and 34.3% respectively) (Petersen *et al*., 2020; Harrison *et al*., 2020).

In contrast, the SARS-CoV-2 that appeared in December, 2019, rapidly triggered a global pandemic. The spread of SARS 2003 ended in June 2003, with 8098 reported cases and 774 deaths, and a fatality rate of 9.7%, with the majority of cases received in hospital. By comparison, the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) - another deadly coronavirus appeared in 2012 and caused 2494 reported cases and 858 deaths in 27 countries, with a fatality rate of 34%. The new SARS-CoV-2 coronavirus is less lethal but more contagious than MERS-CoV or SARS-CoV. The virus was detected in December, 2019 and in 6 months in the first pandemic wave, counting about 10 million infected and 500,000 deaths. Due to the wide clinical spectrum and high transmissibility, the eradication of SARS-CoV-2 with the 2003 SARS-CoV policies and measures does not seem to be a viable goal and equally effective, at least in the short terms (Zhu et al., 2020; Harrison et al., 2020).

Comparative genomics studies show that the genetic characteristics of SARS-CoV-2 and two other coronaviruses with High Fatality Rate (HFR), SARS-CoV and MERS-CoV affect both virulence and pathogenicity, as well as their host-transmitting ability from animals to humans. The evolutive changes of high Case Fatality Rate (CFR) viruses did not happen immediately, but as gradual events and trends. These characteristics are common to high-CFR coronaviruses and their host animals (bats in particular) that infect relatives in the same genus, proving that the emergence of SARS-CoV-2 is a natural process of continuous evolution of the coronavirus and is consistent with the possibility of future zoonotic transmission of other highly pathogenic strains to humans (Abdelrahman *et al*., 2020; Bahrami *et al*., 2020).

From the comparison of these viruses in terms of population-level mortality, transmission and severity pandemic evidence, fatality rate (mortality in individuals with the disease) and population-level mortality it turns out that: a) R0- (basic reproductive rate) of SARS-CoV-2 is similar to or higher than R0 of SARS-CoV and pandemic influenza; b) Mortality due to SARS-CoV-2 and SARS-CoV is prone to people older than 70 years, unlike the influenza pandemics of 1918 and 2009; c) The percentage of symptomatic people seeking hospital treatment is higher for SARS-CoV-2 infections than for the 2009 influenza pandemic; and d) The mortality rate (case fatality rate) is probably about 1%, if asymptomatic individuals / mild symptoms are considered; under investigation and monitoring (SARS-CoV-2 data as of September 2020) (Abdelrahman *et al*., 2020; Toyoshima *et al*., 2020).

SARS-CoV-2: Genetics of human pathogenicity

Current research shows that genetics of specific human (host) genes such as ACE-2, ADAM17 and TMPRSS2, play an important role in susceptibility, pathogenicity and clinical manifestations caused by SARS-CoV-2 in humans. Population-level studies reveal clear negative correlation between the degree of ACE2 expression and the severity of COVID-19 onset. Recent data show that ACE2 is effectively released from human cell membranes through a process that is regulated at different levels and engages two types of membrane proteases: ADAM17 disintegrants and metalloproteinases and TMPRSS2 transmembrane protease. ADAM17 protein acts directly on ACE2 by directing its release into the extracellular space, while TMPRSS2 degrades both ACE2 and SARS-CoV-2 protein S, leading to the fusion of the virus with the membrane and its introduction into the cell. When the proteolytic activity of ADAM17 and TMPRSS2 increases, it triggers releasing rather than cell entry, creating a biochemical barrier to virus infection. This may be because the ACE2-virus interaction occurs away from virus-sensitive tissues. Severity of COVID 19 disease manifestations correlates with low levels of

ACE2 expression, and genetic polymorphisms of ADAM17 expression regulation that affect the intensity of ACE2 extraction from cells (Burgess, 2020; Gussow *et al*., 2020).

Clinical studies have also revealed that incidence and mortality levels from SARS-CoV-2 are significantly different between male and female COVID-19 patients and the disease is associated with pre-existing concomitant conditions, such as cancer and cardiovascular disorders, and in particular individuals with hypertension taking anti-hypertensive medication. The ACE-2 gene is located on the X chromosome. It catalyzes the conversion of angiotensin II to angiotensin- (1-7), which acts as a vasodilator and exerts important modulating effects on the cardiovascular system. TMPRSS2 is an important influencing gene in prostate cancer, identified as an oncogene of the ETS family and detected in several types of tumors (Afewerky 2020; LoPresti *et al*., 2020).

Comparative genomic studies of genetic variants encoding several regions of the ACE2 gene suggest that genomic variants of ACE2 may play an important role in susceptibility to COVID-19 and its associated cardiovascular conditions. In addition to differential polymorphisms which may explain the sensitivity and even outcome in different ethnic populations, the fact that ACE2 is an X-linked gene, might explain the increased / higher risk in males. As such, even in the absence of variations in this gene, the allelic presence of this gene can affect the natural history and prognosis of COVID-19 in males, regardless of ethnicity (Burgess 2020; Di Maria *et al*., 2020).

Comparative genomic studies of coding genetic variants for some regions of the TMPRSS2 gene show that all populations carry p.Val160Met variants, with the highest allele frequency (~ 25%), in Eurasian population with a 40% allele frequency. The p.Asp435Tyr variant is carried only by the European population. These unique widespread polymorphisms in TMPRSS2 provide possible explanations for differential genetic susceptibility to COVID-19 as well as for risk factors, including those with cancer and the high-risk group of patients. Because TMPRSS2 is located at 21q22.3, it can be assumed that individuals with Down syndrome would be at higher risk for COVID-19 infection. Furthermore, the role and oncogenic involvement of TMPRSS2 variants may be related to the results also demonstrated for patients with COVID-19. From RNA analysis it is observed that the expression of TMPRSS2 is higher in ciliated cells and type I alveolar epithelial cells and an increase with aging, which suggests that regulation of TMPRSS2 expression may explain the higher protection of infants and children from COVID-19 (LoPresti *et al*., 2020; Russo *et al*, 2020).

2. CONCLUSIONS

Due to its genetic properties and structure, its very high mutation capacity and the rapid and effective mechanism of infection, SARS-CoV-2 presents a very high degree of pathogenicity expressed in terms of infectivity, trasmissibility of the virus, the human immune response, symptoms and clinical manifestations in patients.

The binding of sars S-protein to ACE2 receptors in various human tissues is the basis of its serious pathological consequences in humans. This also explains the distinctive features of infection of this virus, compared to others of the coronavirus family in terms of infectivity, hospitalization and mortality.

The genetics of SARS-CoV-2 pathogenicity, in addition to the virus genetics, is influenced by a set of human genes involved in the mechanism of infection, severity and response.

Human genes such as ACE-2, ADAM17 and TMPRSS2, play an important role in susceptibility, pathogenicity and clinical manifestations caused by SARS-CoV-2 in humans. ACE2 degree of expression relates to the severity of COVID-19, ADAM17 gene protein acts on ACE2 release from the cell, while TMPRSS2 degrades both ACE2 and SARS-CoV-2 protein S. ACE2 also affects the susceptibility to COVID-19 and the increased incidence in males. The expression of TMPRSS2 in some human cells and its increase with aging, may explain the protection of infants and children from COVID-19.

These results show that the risk and severity of the disease is genetically multifactorial. Further genetic analysis and systematic investigations of functional polymorphisms of potential candidate genes would enable assessment of the most risked and genetically-vulnerable populations and in-depth studies for designing of effective medical treatments and protection for COVID-19.

REFERENCES

Abdelrahman Z**,** Mengyuan Li **M,** Wang **X. 2020.** Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses. *Frontiers in Immunology*, 11:552909.doi: 10.3389/fimmu.2020.552909. eCollection 2020.

Afewerky HK. 2020. Pathology and pathogenicity of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Experimental Biology and Medicine,* **245 (15):** 1299-1307.

Bahrami **A,** Ferns G. 2020. Genetic and pathogenic characterization of SARS-CoV-2. *Future Virology* 15, 8.

Brest P, Refae S, Mograbi B, Hofman P, Milano G. 2020. Host Polymorphisms May Impact SARS-CoV-2 Infectivity. *Trends in Genetics* **36 (11):** 813-815.

Burgess DJ. 2020. Host genetics of coronavirus infection. *Nature Reviews Genetics* 22, 12.

Di Maria E**,** Latini A**,** Borgiani **P,** Novelli G. 2020. Genetic variants of the human host influencing the coronavirus-associated phenotypes (SARS, MERS and COVID-19): rapid systematic review and field synopsis. *Human Genomics* 14, Article number: 30.

Fani **M,** Teimoori **A,** Ghafari **Sh. 2020.** Comparison of the COVID-2019 (SARS-CoV-2) pathogenesis with SARS-CoV and MERS-CoV infections**.** *Future Virology*, 15, 5.

Gussow AB**,** Auslander N**,** Faure G**,** Wolf YI**,** Zhang F**,** Koonin EV. 2020. Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses**.** *PNAS* 117 (26), 15193-15199.

Harrison AG, Lin T, Wang P. 2020. Mechanism of SARS CoV-2 transmission and pathogenesis. *Trends in Immunology*, **40 (12):** 1100-1115.

Hou Y, Zhao J, Martin W, Chung MK, Jehi L, Sharifi N, Erzurum S, Eng Ch, Cheng F. 2020. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. *BMC Medicine*, 18, Article number: 216.

Hu B**,** Guo **H,** Zhou P, Shi **Zh. 2020.** Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, **19:**141–154.

Johns Hopkins Center for Health Security (JHCHS). **2020.** SARS-CoV-2 Genetics. *European Journal of Clinical Microbiology & Infectious Diseases*, **24**: 1–7.

LoPresti M, Beck DB, Duggal P, Cummings DAT, Solomon BD. 2020. The Role of Host Genetic Factors in Coronavirus Susceptibility. *The American Journal of Human Genetics*, **107**: 381-402.

Lotfi M, Rezaei N. 2020. SARS-CoV-2: A comprehensive review from pathogenicity of the virus to clinical consequences. *Journal of Medical Virology*, 6-20.

Ludwig S, Zarbock A. 2020. Coronaviruses and SARS-CoV-2: A Brief Overview. *Anesthesia and Analgesia*, **131(1):**93-96. doi: 10.1213/ANE.0000000000004845.

Nakagawa S, Miyazawa T. 2020. Genome evolution of SARS-CoV-2 and its virological characteristics. *BMC Springer Nature: Inflammation and Regeneration*, 40: Article number: 17.

Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, Storgaard M, l Khalili SA, Simonsen L. 2020. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet/Infection*, 20.

Rabaan AA**,** Al-Ahmed **ShH,** Haque **Sh,** Sah **R,** Tiwari **R, YS,** Dhama **K,** Yatoo MI**,** Aldana DKB**,** Morales AJR. 2020. SARS-CoV-2, SARS-CoV, and MERS-COV: A comparative overview. *Infezioni in Medicina*, **28 (2):**174-184.

Russo R**,** Andolfo **I,** Lasorsa VA**,** Iolascon **A,** Capasso M. **2020**. Genetic Analysis of the Coronavirus SARS-CoV-2 Host Protease TMPRSS2 in Different Populations. *Frontiers in Genetics*, 11, 872.

Sironi M, Hasnain SE, Rosenthal B, Phan T, Fabio Luciani F, Shaw MA, Sallum A, Mirhashemi ME, Morand S, Candelas FG. 2020. SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. *Elsevier Public Health Emergency Collection. Infection, Genetics and Evolution*. 84: 104384.

Smatti MK**,** Al-Sarraj YA**,** Albagha O, Yassine **HM. 2020**. Host Genetic Variants Potentially Associated With SARS-CoV-2: A Multi-Population Analysis. *Frontiers in Genetics*, 02 October 2020. https://doi.org/10.3389/fgene.2020.578523.

Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. 2020. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of Human Genetics*, **65**: 1075–1082.

Tsan T and Lam Y. 2020. Tracking the Genomic Footprints of SARS CoV-2 Transmission. *Trends in genetics*, **36 (8):** 544-546.

Van Dorpa L, Acmana M, Richard D, Shawd L, Forda Ch, Ormonda L, Owena Ch, Panga J, Tana C, Florencia Boshiere AT, Ortiza AT, Ballouxa F. 2020. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution* 83.

Wang H, Li X, Li T, Zhang Sh, Wang L, Wu X, Liu J. 2020. The genetic sequence, origin, and diagnosis of SARS-CoV-2. *European Journal of Clinical Microbiology & Infectious Diseases*, **39(9):**1629-1635. doi: 10.1007/s10096-020-03899-4. Epub 2020 Apr 24.

Zeng H**,** Dichio V**,** Horta ER**,** Thorell K**,** Aurell E. 2020. Global analysis of more than 50,000 SARS-CoV-2 genomes reveals epistasis between eight viral genes. *PNAS* 117 (49), 31519-31526.

Zhu Zh**,** Lian **X,** Su **X,** Wu **W,** Marraro GA, Zeng **Y. 2020.** From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Springer Nature. Respiratory Research*, 21, Article number: 224