

The genetic landscape of COVID-19: A South Asian perspective

Vineeth Thomas,¹ Jennifer Audsley,² Nitin Kapoor³

Abstract

COVID-19 has taken the world by storm in the ongoing pandemic. The virus responsible for COVID-19 disease is 'severe acute respiratory syndrome coronavirus-2' SARS-CoV-2, an enveloped RNA beta-coronavirus from the family Coronaviridae. There have been similar beta-coronavirus disease outbreaks previously: Severe acute respiratory syndrome (SARS - 2002) and Middle East respiratory syndrome (MERS - 2012) epidemics. SARS-CoV-2 origins have been traced to bat reservoirs. A virus with a high capacity for mutation, SARS-CoV-2 poses unique challenges both in the current form of disease control and management, while also leaving the door open for future novel diseases and pandemics. An understanding of the virion structure and genomic organisation will help us in understanding their origins and likely course of future evolution. Moreover, novel cost-effective methodologies for genetic surveillance may help in mitigating the emergence of these viral infections in future. In this manuscript, the authors have detailed the unique aspects of the SARS-CoV-2 virus genome and its clinical implications.

Keywords: COVID-19, Genetic mutations, Next generation sequencing, SARS-CoV-2, South Asian.

DOI: <https://doi.org/10.5455/JPMA.09>

Introduction

Viruses are obligate intracellular, microorganisms that vary in size from 10-300nm. By comparison, the size of bacteria is 1000nm and red blood cells are 10,000nm. They carry their own genetic material, either DNA or RNA, but never both. This coupled with the lack of a protein-

synthesizing apparatus makes them completely dependent on host cellular machinery for genomic replication and virion synthesis.

Following the first reports of pneumonia of unknown aetiology in Wuhan in the Hubei province of China in December 2019, the causative agent identified was provisionally named novel coronavirus 2019 (nCoV-19).¹ The outbreak was declared a pandemic by the World Health Organisation on 11th March 2020, and the disease has been named COVID-19. nCoV-19 was later classified and named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the Coronaviridae Study Group of the International Taxonomy of Viruses.²

Coronaviruses are enveloped RNA viruses of the Nidovirales order (Figure-1). The genus betacoronavirus includes, amongst others, SARS-CoV, MERS-CoV and now SARS-CoV-2. Externally, they are characterised by a spherical shape ~125nm diameter with club-like spikes over its surface and a nucleocapsid which is ~10nm in diameter. The virus has a non-segmented, positive-sense RNA genome (~30 kb) with a 5' cap and a 3' tail, which allows it to act as messenger RNA (mRNA) for the replicase enzyme in the host cell. Two-thirds (~20 kb) of the viral genome codes for non-structural proteins while one-third (~10 Kb) codes for the structural proteins.³

Virion structure

Coronavirus virions have club-like spikes originating from the surface of the envelope giving them the characteristic appearance of a solar corona (Figure-2). Within this envelope are helically symmetric nucleocapsids. There are 4 main structural proteins which are encoded within the 3' end of the viral genome:

1. Surface glycoprotein (S)
2. Membrane (M)
3. Envelope (E)
4. Nucleocapsid (N)
5. Haemagglutinin-esterase (HE)

.....
¹Department of Internal Medicine, Christian Medical College, Vellore (TN)-632004, India, ²The Peter Doherty Institute for Infection and Immunity, The University of Melbourne and Royal Melbourne Hospital, Melbourne, Victoria, Australia, ³Department of Endocrinology, Diabetes and Metabolism, Christian Medical College, Vellore (TN) -632004, India, and Non Communicable Disease Unit, Melbourne School of Population and Global Health, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Victoria, Australia.

Correspondence: Nitin Kapoor. Email: nitin.kapoor@cmcvellore.ac.in

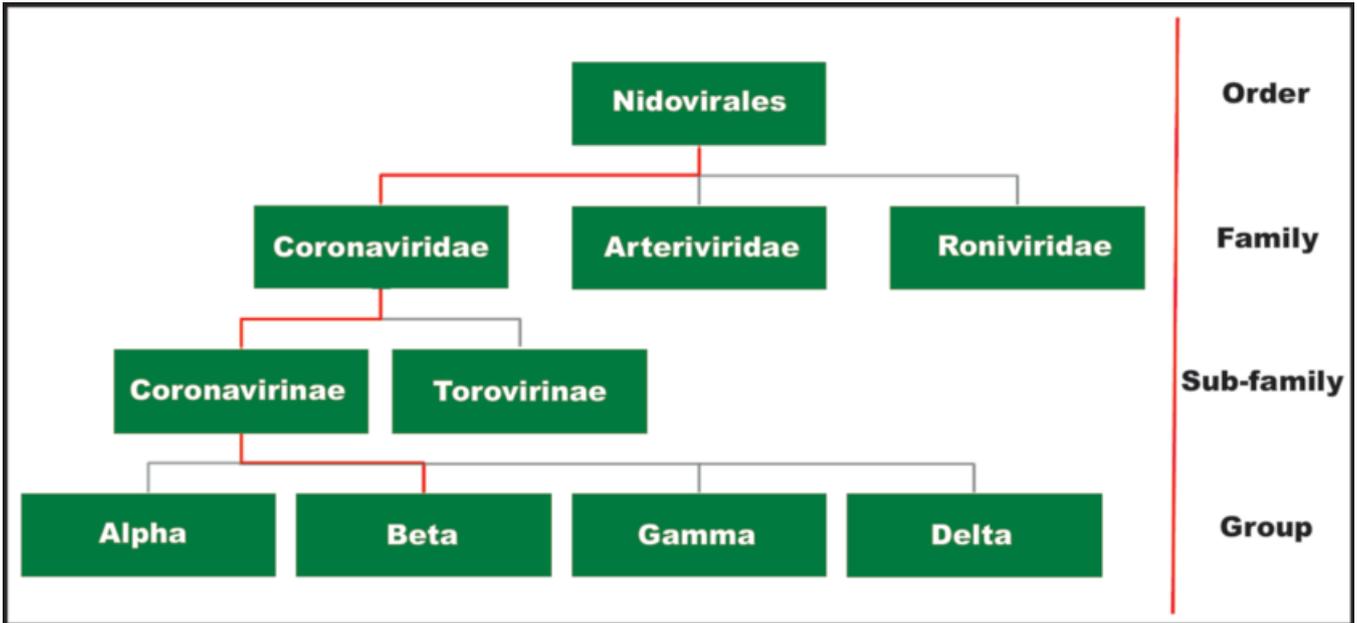


Figure-1: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) taxonomy. Adapted from Fehr et al.²

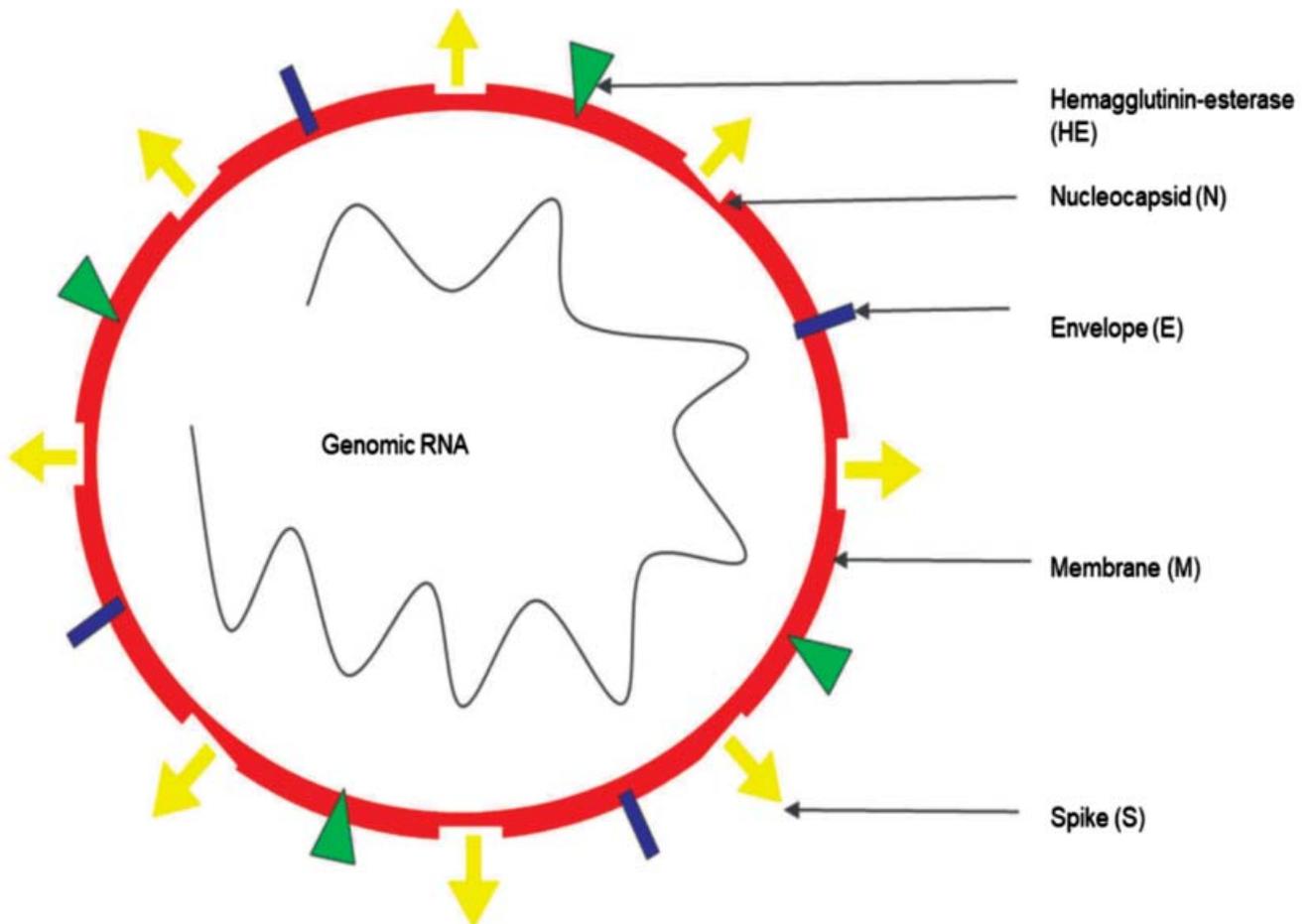


Figure-2: The structure of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Adapted from Shereen et al.²⁵

The main function of the S-protein is to mediate the attachment of the virus to the host cell while the M-protein gives shape to the virion and is the most abundant structural protein. The E-protein is found in small quantities and facilitates the assembly and release of the virus. The N-protein is the only protein in the nucleocapsid and is heavily phosphorylated. These proteins bind to the viral genome and provide its helical pattern. They are involved in the unpacking of the genome during viral replication and subsequently packaging into viral particles. The haemagglutinin-esterase enzyme present on the capsid is present only in the beta-subgroup of coronavirus and its activity is to enhance S protein-mediated cell entry.⁴

Genomic Organisation

The genome of coronavirus encodes for two sets of proteins, (i) the non-structural proteins which accounts for ~20kb of the genome, and (ii) the structural proteins with accessory proteins (~10kb).

The organisation of structural proteins within the genome is interspersed with accessory proteins (Figure-3). The structural proteins encoded are responsible for the final translation into the surface glycoprotein, membrane, envelope, nucleocapsid, and haemagglutinin receptor.⁵

Human Coronavirus

Human coronaviruses are generally responsible for mild

infections and account for around 25% of all viral illnesses. Most infections are self-limiting upper respiratory tract infections, however, in the high-risk age groups of elderly or individuals with underlying medical comorbidities, they tend to cause severe respiratory illness.⁶ There have previously been two beta-coronaviruses that resulted in a major global public health crisis, as listed below:^{7,8}

- a) Severe Acute Respiratory Distress Syndrome (SARS):
 - Classified as group-2b, beta-coronavirus
 - Original reservoir bats with Himalayan civets as intermediate host
 - Virion affinity to angiotensin converting enzyme 2 (ACE2) receptor
- b) Middle East Respiratory Syndrome (MERS):
 - Classified as group 2c, beta-coronavirus
 - Original reservoir bats with camels as intermediate host
 - Virion affinity to dipeptidyl peptidase-4 (DPP-4) receptor

Key features of SARS, MERS and COVID-19 are listed in Table-1.^{9,10} The factors essential for disease emergence are represented in Figure-4.

High Mutant Variations in Coronavirus

One of the important reasons why some coronaviruses

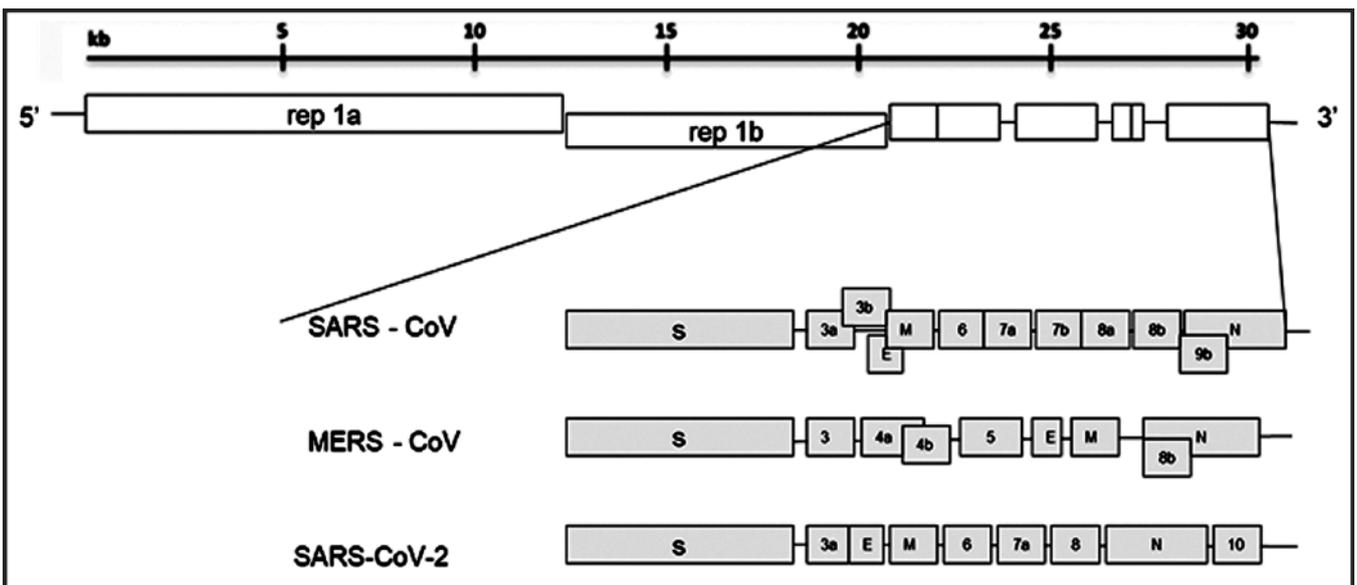


Figure-3: The genomic comparison of different coronaviruses. Adapted from Mousavizadeh et al[19 and Van Boheemen et al.²⁶



Figure-4: Essential factors for disease emergence. Adapted from Morse SS.²⁷

Table-1: Comparison of Severe acute respiratory syndrome (SARS), Middle East Respiratory Syndrome (MERS), and coronavirus disease (COVID-19). Adapted from Khafaie et al.²⁸

	Original reservoir host	New intermediary Host	Receptor	Fatality Rate
SARS (2002)	Bats	Himalayan civets	ACE-2 receptor	10%
MERS (2012)	Bats	Camels	DPP-4 receptor	34%
COVID- 19 (2019)	Bats	Pangolin*	ACE-2 receptor	3.4%

* likely intermediary host, yet to be confirmed definitively.

cross over to other species is the high mutation rates. Viral mutations can occur by three mechanisms:

- (i) Natural change in nucleic acid bases that constitute genetic material
- (ii) Physical mutagens on genetic material (X-rays, UV light)
- (iii) Mistakes in proofreading of genetic material during replication

In contrast to DNA viruses or eukaryotic cells, RNA viruses lack DNA polymerase which is involved in proofreading of replication enzymes. As an illustration, DNA viruses have a mutation rate of the order 10⁻⁸ to 10⁻⁶ while RNA viruses have a higher mutation rate of 10⁻⁶ to 10⁻⁴.¹¹ This high replication rate results in a high mutation rate in coronaviruses. The error rate for DNA viruses is estimated to be 10⁻⁸ to 10⁻¹¹ errors per incorporated nucleotide and in RNA viruses this number increases to 10⁻³ to 10⁻⁴ errors per incorporated nucleotide, suggesting that RNA viruses generate mutants frequently, perhaps even once per replication cycle.¹²

Coronaviruses in non-human reservoirs must undergo multiple genetic mutations before they are able to become infective to humans. However, this is not a frequent occurrence because mutations that occur do not always persist in the population especially when they negatively impact the survival of the virus (viral fitness). There are occasionally mutations which improve viral fitness and accumulation of such mutations may result in

cross-species transmission, culminating in disease or as asymptomatic carriage in the host helping to propagate their spread.

As a result of this high mutant variation and replication rate, the population of viruses consist of a large group with related genotypes called 'quasispecies'.¹³ Previous viral quasispecies studies have found that viruses with more diverse mutations are more likely to thrive in adverse growth conditions than their individual variant populations. As of 5th April, GenBank, an organisation that shares genetic sequence data of viruses including SARS-CoV-2 had documented 446 genome sequences isolated from patients from around the globe.¹⁴

With closely related bat coronaviruses having 96.2% sequence similarity to the human SARS-CoV2 virus, the original reservoir of SARS-CoV-2 is most likely bats, which are known reservoirs of multiple viruses.¹⁵ Prior to the current SARS-CoV2 outbreak, genetically similar viruses were identified in bats.^{16,17} As with SARS and MERS, an intermediate host was required to transmit the infection to humans as the bats ecological system rarely intersects with humans. Pangolins (scaly anteaters) are mammals commonly trafficked in wet markets and coronavirus isolates from Malayan pangolins were found to have 97% similarity with SARS-CoV-2.¹⁸ The genome structure of SARS-CoV-2 is similar to other beta-coronaviruses (Figure 3), with the gene order 5' replicase-S-E-M-N-3'. SARS and MERS virus have a 72%

and 50% similarity with respect to their viral genome.¹⁹ Minor structural variations of SARS-CoV-2 in relation to previously documented coronavirus genomes appear to account for its infectivity. For example, of SARS-CoV-2 contains a furin-like cleavage site at the junction of S1 and S2 on the S-protein which is thought to result in higher infectivity rate as compared to previous beta-coronaviruses.²⁰ This S-protein allows the virus to enter cells via the angiotensin-converting enzyme 2 (ACE2) receptor and uses the serine protease enzyme TMPRSS2 to complete this process.²¹ More research is required to clearly understand the evolution of this virus and its transmissibility.

As infection spreads globally, the accumulation of mutations has been observed in the SARS-CoV-2 genome. A recently published phylogenetic network analysis of 160 complete SARS-CoV-2 genomes identified three central variants (A, B and C) which could be differentiated by amino acid changes.²² The isolates were from patients in China, East Asia, USA, Canada, Europe, Australia, Mexico and Brazil. Type A is the ancestral type closest to bat coronavirus. Type B differs from the ancestral A by two mutations (T8782C and C28144T) and is the most common type in East Asia. Type C differs from its parent Type B by the mutation G26144T and was absent in the mainland Chinese samples examined, but found in samples from Europe, Singapore, Hong Kong, Taiwan, and South Korea. Types A and C are found in significant proportions outside East Asia.

As COVID-19 pandemic is rapidly evolves, future policy decisions on its active genetic surveillance should be considered. This may help in predicting and mitigating the emergence of these viral infections in humans. Given their high genetic diversity, a capture-based NGS (next generation sequencing) approach using baits has recently been used as a preferred method for virus discovery to cover unbiased sequencing of bat coronaviruses.²³ This approach has been used in cost-effective diagnosis in South Asian low middle income countries (LMIC) for other genetic disorders.²⁴ This would be especially relevant in many LMIC, as it may significantly reduce cost and enhances the sensitivity of these surveillance studies.

Acknowledgement: We would like to acknowledge the ENCORE (Excellence in Non-Communicable disease Research (ENCORE) programme (between Australian and India) for facilitating this joint publication between the two countries.

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020;382:1199-1207. doi: 10.1056/NEJMoa2001316.
2. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020;5:536-544. doi: 10.1038/s41564-020-0695-z.
3. Bárcena M, Oostergetel GT, Bartelink W, Faas FGA, Verkleij A, Rottier PJM, et al. Cryo-electron tomography of mouse hepatitis virus: insights into the structure of the coronavirus. *Proc Natl Acad Sci U S A* 2009;106:582-7.
4. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol Clifton NJ* 2015;1282:1-23.
5. Brian DA, Baric RS. Coronavirus genome structure and replication. *Curr Top Microbiol Immunol* 2005;287:1-30.
6. Paules CI, Marston HD, Fauci AS. Coronaviruses - more than just the common cold. *JAMA* 2020. doi: 10.1001/jama.2020.0757.
7. Yang X-L, Hu B, Wang B, Wang M-N, Zhang Q, Zhang W, et al. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. *J Virol* 2015;90:3253-6.
8. Memish ZA, Cotten M, Meyer B, Watson SJ, Alsaifi AJ, Al Rabeeah AA, et al. Human infection with MERS coronavirus after exposure to infected camels, Saudi Arabia, 2013. *Emerg Infect Dis* 2014;20:1012-5.
9. Choudhry H, Bakhrebah MA, Abdulaal WH, Zamzami MA, Baothman OA, Hassan MA, et al. Middle East respiratory syndrome: pathogenesis and therapeutic developments. *Future Virol* 2019;14:237-46.
10. Gu J, Korteweg C. Pathology and pathogenesis of severe acute respiratory syndrome. *Am J Pathol* 2007;170:1136-47.
11. Peck KM, Lauring AS. Complexities of viral mutation rates. *J Virol* 2018;92:e01031-17. doi: 10.1128/JVI.01031-17.
12. Fleischmann WR. Viral Genetics. In: Baron S, editor. *Medical Microbiology* [Internet]. 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston; 1996 [cited 2020 Apr 7]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK8439/>
13. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 2006;439:344-8.
14. SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) Sequences [Internet] [cited 2020 Apr 7]. Available from: <https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>
15. Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev* 2006;19:531-45.
16. Hu B, Zeng L-P, Yang X-L, Ge X-Y, Zhang W, Li B, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog* 2017;13:e1006698. doi: 10.1371/journal.ppat.1006698.
17. Ge X-Y, Li J-L, Yang X-L, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 2013;503:535-8.
18. Lam TT-Y, Shum MH-H, Zhu H-C, Tong Y-G, Ni X-B, Liao Y-S, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* 2020. doi: 10.1038/s41586-020-2169-0.
19. Genotype and phenotype of COVID-19: Their roles in pathogenesis [Internet] [cited 2020 Apr 14]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7138183/>
20. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade.

- Antiviral Res 2020;176:104742. doi: 10.1016/j.antiviral.2020.104742.
21. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020;181:271-280.e8. doi: 10.1016/j.cell.2020.02.052
 22. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci U S A* 2020;117:9241-9243. doi: 10.1073/pnas.2004999117.
 23. Li B, Si HR, Zhu Y, Yang XL, Anderson DE, Shi ZL, et al. Discovery of bat coronaviruses through surveillance and probe capture-based Next-Generation Sequencing. *mSphere* 2020;5:e00807-19. doi: 10.1128/mSphere.00807-19.
 24. Kapoor N, Chapla A, Furler J, Paul TV, Harrap S, Oldenburg B, et al. Genetics of obesity in consanguineous populations - A road map to provide novel insights in the molecular basis and management of obesity. *EBioMedicine* 2019;40:33-4.
 25. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. *J Adv Res* 2020;24:91-8.
 26. van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, et al. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio* 2012;3:e00473-12. doi: 10.1128/mBio.00473-12.
 27. Morse SS. Factors in the emergence of infectious diseases. *Emerg Infect Dis* 1995;1:7-15.
 28. Khafaie MA, Rahim F. Cross-country comparison of case fatality rates of COVID-19/SARS-COV-2. *Osong Public Health Res Perspect* 2020;11:74-80.
-