Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Microbiology, Immunology and Infection xxx (xxxx) xxx

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jmii.com

Review Article

Genotype and phenotype of COVID-19: Their roles in pathogenesis

Leila Mousavizadeh [a], **Sorayya Ghasemi** [b,c,*]

[a] *Department of Virus-host Interaction, Heinrich-Pette-Institut (HPI), Martinistrasse 52, 20251 Hamburg, Germany*
[b] *Cellular and Molecular Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran*
[c] *Cancer Research Center, Shahrekord University of Medical Sciences, Shahrekord, Iran*

Received 20 March 2020; received in revised form 22 March 2020; accepted 22 March 2020
Available online ■ ■ ■

KEYWORDS
COVID-19;
Genotype;
Phenotype;
Pathogenesis

Abstract COVID-19 is a novel coronavirus with an outbreak of unusual viral pneumonia in Wuhan, China, and then pandemic. Based on its phylogenetic relationships and genomic structures the COVID-19 belongs to genera Betacoronavirus. Human Betacoronaviruses (SARS-CoV-2, SARS-CoV, and MERS-CoV) have many similarities, but also have differences in their genomic and phenotypic structure that can influence their pathogenesis. COVID-19 is containing single-stranded (positive-sense) RNA associated with a nucleoprotein within a capsid comprised of matrix protein. A typical CoV contains at least six ORFs in its genome. All the structural and accessory proteins are translated from the sgRNAs of CoVs. Four main structural proteins are encoded by ORFs 10, 11 on the one-third of the genome near the 3′-terminus. The genetic and phenotypic structure of COVID-19 in pathogenesis is important. This article highlights the most important of these features compared to other Betacoronaviruses.

Introduction

Coronaviruses are involved in human and vertebrate's diseases.[1] Coronaviruses are members of the subfamily Coronavirinae in the family Coronaviridae and the order Nidovirales. The recent emergence of a novel coronavirus with an outbreak of unusual viral pneumonia in Wuhan, China and then pandemic outbreak is 2019-nCoV or COVID-19. Based on its phylogenetic relationships and genomic

* Corresponding author. Cellular and Molecular Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Rahmatiyeh, Shahrekord, 8813833435, Iran. Fax: +98 3833330709.
E-mail addresses: lmosavi2007@yahoo.com (L. Mousavizadeh), sorayya.ghasemi@gmail.com (S. Ghasemi).

https://doi.org/10.1016/j.jmii.2020.03.022

Please cite this article as: Mousavizadeh L, Ghasemi S, Genotype and phenotype of COVID-19: Their roles in pathogenesis, Journal of Microbiology, Immunology and Infection, https://doi.org/10.1016/j.jmii.2020.03.022

structures the COVID-19 belongs to genera Betacoronavirus which has a close similarity of the sequences of COVID19 to that of severe acute respiratory syndrome-related coronaviruses (SARSr-CoV) and the virus uses ACE2 as the entry receptor-like SARS-CoV.[2] These similarities of the SARS-CoV-2 to the one that caused the SARS outbreak (SARS-CoVs) the Coronavirus Study Group of the International Committee on Taxonomy of Viruses termed the virus as SARS-CoV-2.[3] The understanding of the genetic and phenotypic structure of COVID-19 in pathogenesis is important for the production of drugs and vaccines. So, in this review article, we provide the newest genetic and phenotype features of COVID-19 to investigate the role of these two factors in the pathogenesis and comparing it with its families.

Coronavirus genome structure and life cycle

COVID-19 is a spherical or pleomorphic enveloped particles containing single-stranded (positive-sense) RNA associated with a nucleoprotein within a capsid comprised of matrix protein. The envelope bears club-shaped glycoprotein projections. Some coronaviruses also contain a hem agglutinin-esterase protein (HE)[4] (Fig. 1).

Coronaviruses possess the largest genomes (26.4—31.7 kb) among all known RNA viruses, with G + C contents varying from 32% to 43%. Variable numbers of small ORFs are present between the various conserved genes (ORF1ab, spike, envelope, membrane and nucleocapsid) and, downstream to the nucleocapsid gene in different coronavirus lineages. The viral genome contains distinctive features, including a unique N-terminal fragment within the spike protein. Genes for the major structural proteins in all coronaviruses occur in the 5′—3′ order as S, E, M, and N[5](Fig. 2).

A typical CoV contains at least six ORFs in its genome. Except for Gammacoronavirus that lakes nsp1, the first ORFs (ORF1a/b), about two-thirds of the whole genome length, encode 16 nsps (nsp1-16). ORF1a and ORF1b contain a frameshift in between which produces two polypeptides:

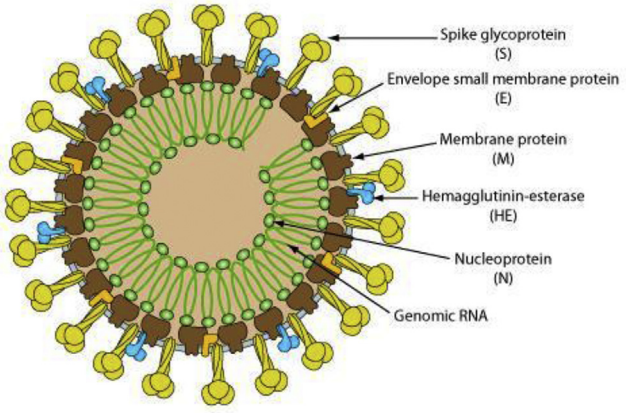

Figure 1. Schematic of a coronavirus — this new virus probably looks a lot like this. From Biowiki (http://ruleof6ix.fieldofscience.com/2012/09/a-new-coronavirus-should-you-care.html).

pp1a and pp1ab. These polypeptides are processed by virally encoded chymotrypsin-like protease (3CLpro) or main protease (Mpro) and one or two papain-like protease into 16 nsps. All the structural and accessory proteins are translated from the sgRNAs of CoVs. Four main structural proteins contain spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins are encoded by ORFs 10, 11 on the one-third of the genome near the 3′-terminus.[6,7] Besides these four main structural proteins, different CoVs encode special structural and accessory proteins, such as HE protein, 3a/b protein, and 4a/b protein (Fig. 2B, lower panel). These mature proteins are responsible for several important functions in genome maintenance and virus replication.[6]

There are three or four viral proteins in the coronavirus membrane. The most abundant structural protein is the membrane (M) glycoprotein; it spans the membrane bilayer three times, leaving a short NH2-terminal domain outside the virus and a long COOH terminus (cytoplasmic domain) inside the virion.[4] The spike protein (S) as a type I membrane glycoprotein constitutes the peplomers. In fact, the main inducer of neutralizing antibodies is S protein. Between the envelope proteins with exist a molecular interaction that probably determines the formation and composition of the coronaviral membrane. M plays a predominant role in the intracellular formation of virus particles without requiring S. In the presence of tunicamycin coronavirus grows and produces spikeless, noninfectious virions that contain M but devoid of S.[4,5]

Comparison of SARS-CoV2 (COVID-19), SARS-CoV, and MERS-CoV

The 5′UTR and 3′UTR are involved in inter- and intramolecular interactions and are functionally important for RNA—RNA interactions and for binding of viral and cellular proteins.[8] At 5 end, Pb1ab is the first ORF of the whole genome length encoding non-structural proteins with size of 29844bp (7096aa), 29751bp (7073aa) and 30119bp (7078) in COVID-19, SARS-CoV; and MERS-CoV, respectively. Even with comparison of the spike protein at 3′ end, among the coronaviruses specifically these three betacoronaviruses, the difference was visualized, 1273aa, 21493aa, and 1270aa in COVID-19, SARS-CoV, and MERS-CoV, respectively. Genetically, COVID-19 was less similar to SARS-CoV (about 79%) and MERS-CoV (about 50%). The arrangement of nucleocapsid protein (N), envelope protein(E), and membrane protein (M) among betacoronaviruses are different as depicted in Fig. 3.[9]

The role of replication process in pathogenicity

SARS-CoV-2 (COVID-19) binds to ACE2 (the angiotensin-converting enzyme 2) by its Spike and allows COVID-19 to enter and infect cells. In order for the virus to complete entry into the cell following this initial process, the spike protein has to be primed by an enzyme called a protease. Similar to SARS-CoV, SARS-CoV-2 (COVID-19) uses a protease called TMPRSS2 to complete this process.[10,11] In order to

Please cite this article as: Mousavizadeh L, Ghasemi S, Genotype and phenotype of COVID-19: Their roles in pathogenesis, Journal of Microbiology, Immunology and Infection, https://doi.org/10.1016/j.jmii.2020.03.022

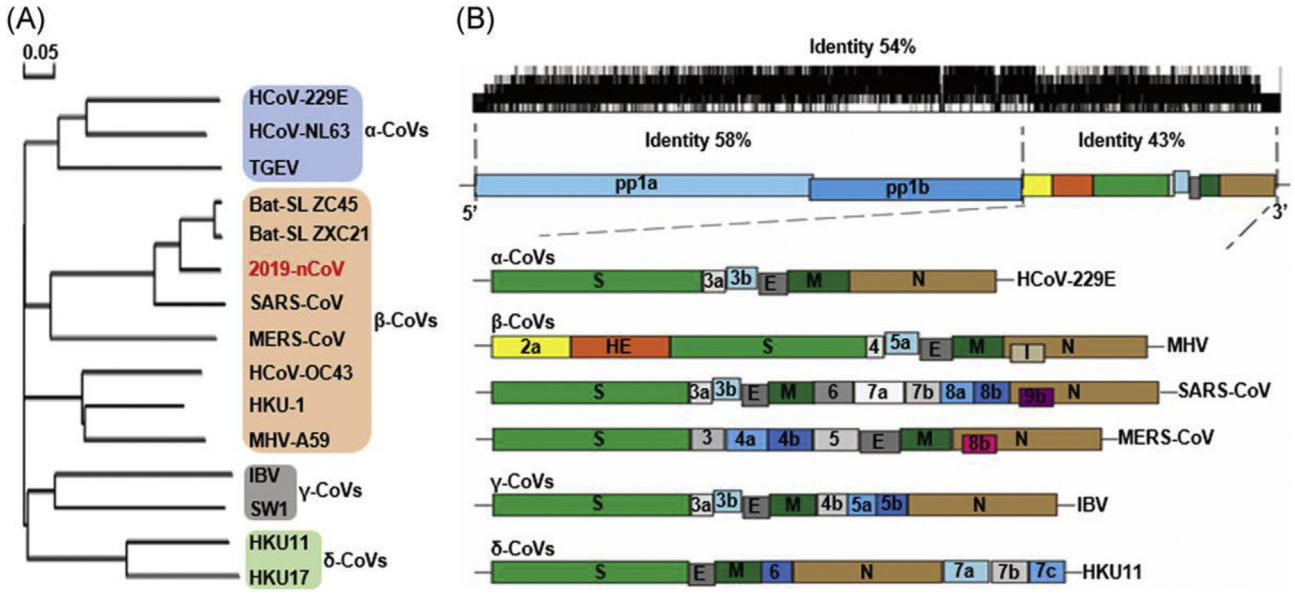

Figure 2. The genomic structure and phylogenetic tree of coronaviruses: A, the phylogenetic tree of representative CoVs, with the new coronavirus COVID-19 shown in red. B, The genome structure of four genera of coronaviruses: two long polypeptides 16 nonstructural proteins have proceeded from Pp1a and pp1b represent. S, E, M, and N are represented of the four structural proteins spike, envelope, membrane, and nucleocapsid. COVID-19; CoVs, coronavirus; HE, hemagglutinin-esterase. Viral names: HKU, coronaviruses identified by Hong Kong University; HCoV, human coronavirus; IBV, infectious bronchitis virus; MHV, murine hepatitis virus; TGEV, transmissible gastroenteritis virus.[1]

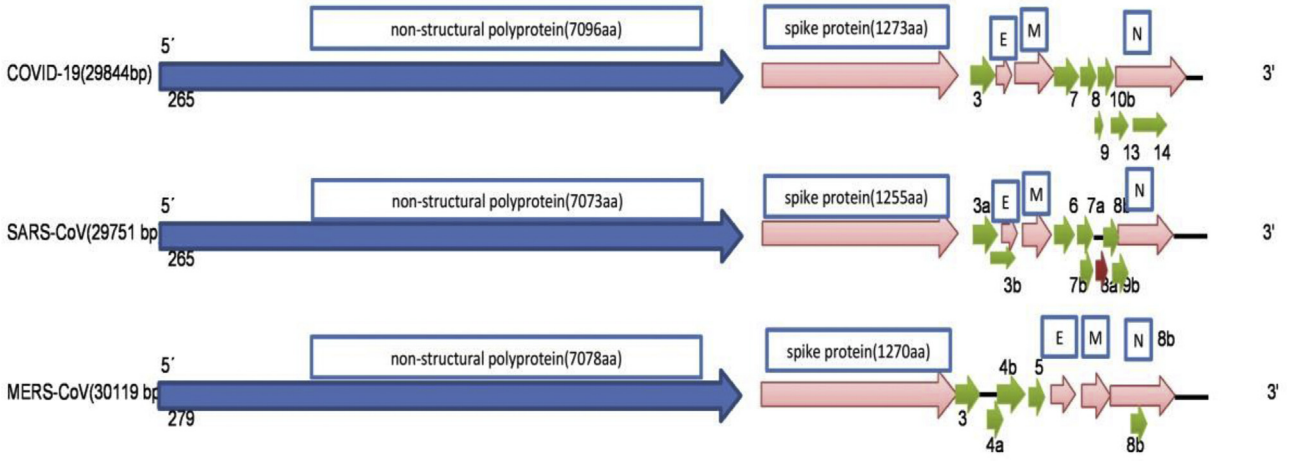

Figure 3. The 5′ UTR and 3′ UTR and coding region of COVID-19, SARS-CoV, and MERS-CoV. The numbers of base pairs among betacoronaviruses are shown. This figure is modified from the sequence comparison and genomic organization of 2019-nCoV, 2020.[9] The differences in the arrangement of the envelope (E), membrane (M), and nucleoprotein (N) among COVID-19, SARS-CoV, and MERS-CoV are shown at 3′ end.

attach virus receptor (spike protein) to its cellular ligand (ACE2), activation by TMPRSS2 as a protease is needed (Fig. 4).[10]

After the virus enters the host cell and uncoats, the genome is transcribed and then translated. Coronavirus genome replication and transcription takes place at cytoplasmic membranes and involve coordinated processes of both continuous and discontinuous RNA synthesis that are mediated by the viral replicate, a huge protein complex encoded by the 20-kb replicase gene.[12] The replicase complex is believed to be comprised of up to 16 viral subunits and a number of cellular proteins. Besides RNA-dependent RNA polymerase, RNA helicase, and protease activities, which are common to RNA viruses, the coronavirus replicase was recently predicted to employ a variety of RNA processing enzymes that are not (or extremely rarely) found in other RNA viruses and include putative sequence-specific endoribonuclease, 3′-to-5′ exoribonuclease, 2′-*O*-ribose methyltransferase, ADP ribose 1′-phosphatase and, in a subset of group 2 coronaviruses, cyclic

Please cite this article as: Mousavizadeh L, Ghasemi S, Genotype and phenotype of COVID-19: Their roles in pathogenesis, Journal of Microbiology, Immunology and Infection, https://doi.org/10.1016/j.jmii.2020.03.022

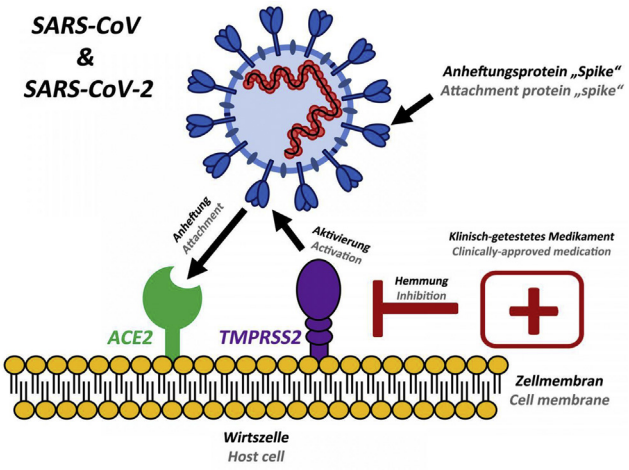

Figure 4. The attachment protein "spike" of the new coronavirus COVID-19 andSARS-CoV use the same cellular attachment factor (ACE2) and the cellular protease TMPRSS2 for their activation. Existing, clinically approved drugs directed against TMPRSS2 inhibit SARS-CoV-2 infection of lung cells.[10]

phosphodiesterase activities.[13,14] The proteins are assembled at the cell membrane and genomic RNA is incorporated as the mature particle forms by budding from the internal cell membranes.[15]

Factors affecting virus pathogenesis

Co-morbidities are cardiovascular and cerebrovascular disease as well as diabetes. Several abnormalities also have been observed including cellular immune deficiency, coagulation activation, myocardia injury, hepatic and kidney injury, and secondary bacterial infection.[16] In the majority of cases of severe disease and death, lymphopenia and sustained inflammation have been recorded. Notably, these observations in COVID-19 patients are similar to those who suffered from severe acute respiratory syndrome (SARS) during the 2003 epidemic. There may be a biological mechanism behind this epidemiological anomaly.[17]

Several kinds of vaccines and antiviral drugs that are based on S protein have been previously evaluated. Du et al. showed vaccines can be based on the S protein include full-length S protein, viral vector, DNA, recombinant S protein and recombinant RBD protein. Considering that, in the in vitro study, antiviral therapies are design based on S protein include RBD—ACE2 blockers, S cleavage inhibitors, fusion core blockers, neutralizing antibodies, protease inhibitors, S protein inhibitors, and small interfering RNAs.[18] There are some recombinant compounds such as IFN with ribaverin which has only limited effects against COVID-19 infection.[1] The receptor-binding domain of SARS-CoV-2 has a higher affinity for ACE2, while it is a lower affinity for SARS-CoV.[1,19] Angiotensin-converting enzyme (ACE) and its homologue ACE2, belongs to the ACE family of dipeptidylcarboxy dipeptidase. However, their physiological functions are varied. On the other hand, ACE2 serves as the binding site for COVID-19. Based on this information, Gurwitz suggested using available angiotensin receptor 1 (AT1R) blockers, such as losartan, as therapeutics for reducing the severity of COVID-19 infections.[20] At present therapy is based on identifying and developing monoclonal antibodies that are specific and effective against COVID-19 combines with remdesivir as a novel nucleotide analog prodrug that was used for the treatment of the Ebola virus disease.[17,21] To understand the rate of virus spread among people, it is crucial to figure out whether COVID-19 is mutating to improve its binding to human receptors for infection considering its high mutation rate. Any adaptation in the COVID-19 sequence that might make it more efficient at transmitting among people might also boost its virulence.[22] The COVID-19 is expected to become less virulent through human to human transmissions due to genetic bottlenecks for RNA viruses often occur during respiratory droplet transmissions.[3]

Conclusion

At present, there is no specific treatment for COVID-19. Given the high rate of transmission of this virus between humans and its pandemics, it is important to identify the basis of its replication, structure, and pathogenicity for discovering a way to the special treatment or the prevention. Due to the high similarity of the virus to its families, efforts have been made to provide medicines and vaccines for COVID-19. Differences in the length of the spike as it is longer in COVID-19 are likely to play an important role in the pathogenesis and treatment of this virus. However, identifying the specific molecular details of the virus is helpful in achieving treatment goals.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare no potential conflicts of interest to disclose.

Acknowledgments

The authors would like to thank the Shahrekord University of Medical Sciences.

References

1. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol* 2020;**92**:418—23.
2. Seah I, Agrawal R. Can the coronavirus disease 2019 (COVID-19) affect the eyes? A review of coronaviruses and ocular implications in humans and animals. *Ocul Immunol Inflamm* 2020:1—5.
3. Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. *Features, evaluation and treatment coronavirus (COVID-19).* StatPearls [Internet]. StatPearls Publishing; 2020.

4. de Haan CAM, Kuo L, Masters PS, Vennema H, Rottier PJM. Coronavirus particle assembly: primary structure requirements of the membrane protein. *J Virol* 1998;**72**(8):6838—50.
5. Woo PCY, Huang Y, Lau SKP, Yuen K-Y. Coronavirus genomics and bioinformatics analysis. *Viruses* 2010;**2**(8):1804—20.
6. 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**(6). e00473—12.
7. Czub M, Weingartl H, Czub S, He R, Cao J. Evaluation of modified vaccinia virus Ankara based recombinant SARS vaccine in ferrets. *Vaccine* 2005;**23**(17—18):2273—9.
8. Yang D, Leibowitz JL. The structure and functions of coronavirus genomic 3′ and 5′ ends. *Virus Res* 2015;**206**:120—33.
9. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;**395**(10224):565—74.
10. 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. https://doi.org/10.1016/j.cell.2020.02.052.
11. Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak—an update on the status. *Mil Med Res* 2020;**7**(1):1—10.
12. Sola I, Almazan F, Zuniga S, Enjuanes L. Continuous and discontinuous RNA synthesis in coronaviruses. *Ann Rev Virol* 2015;**2**:265—88.
13. Ziebuhr J. *The coronavirus replicase. Coronavirus replication and reverse genetics*. Springer; 2005. p. 57—94.
14. Almazán F, DeDiego ML, Galán C, Escors D, Álvarez E, Ortego J, et al. Construction of a severe acute respiratory syndrome coronavirus infectious cDNA clone and a replicon to study coronavirus RNA synthesis. *J Virol* 2006;**80**(21):10900—6.
15. McIntosh K, Peiris JSM. *Coronaviruses. Clinical virology*. 3rd ed. American Society of Microbiology; 2009. p. 1155—71.
16. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;**395**(10223):507—13.
17. Liu C, Zhou Q, Li Y, Garner LV, Watkins SP, Carter LJ, et al. *Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases*. ACS Publications; 2020.
18. Du L, He Y, Zhou Y, Liu S, Zheng B-J, Jiang S. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat Rev Microbiol* 2009;**7**(3):226—36.
19. Prabakaran P, Xiao X, Dimitrov DS. A model of the ACE2 structure and function as a SARS-CoV receptor. *Biochem Biophys Res Commun* 2004;**314**(1):235—41.
20. Gurwitz D. Angiotensin receptor blockers as tentative SARS-CoV-2 therapeutics. *Drug Dev Res* 2020:1—4. https://doi.org/10.1002/ddr.21656.
21. Sheahan TP, Sims AC, Graham RL, Menachery VD, Gralinski LE, Case JB, et al. Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci Transl Med* 2017;**9**(396).
22. Martinez MA. *Compounds with therapeutic potential against novel respiratory 2019 coronavirus. Antimicrobial Agents and Chemotherapy*. 2020.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmii.2020.03.022.

Please cite this article as: Mousavizadeh L, Ghasemi S, Genotype and phenotype of COVID-19: Their roles in pathogenesis, Journal of Microbiology, Immunology and Infection, https://doi.org/10.1016/j.jmii.2020.03.022