

Journal of Hematology and Allied Sciences

Journal of
Hematology and Allied
Sciences (JHAS) $\overline{\mathbf{Q}}$

What the Expert Says Coronavirus-2019: Origin and evolution

Nitai Chandra Mandal

Department of Biochemistry, Bose Institute, P-1/12, CIT Scheme VII, Kolkata-700054, West Bengal, India

ABSTRACT

***Corresponding author:**

Nitai Chandra Mandal Department of Biochemistry, Bose Institute, P-1/12, CIT Scheme VII, Kolkata-700054, West Bengal, India

mandalnc2003@yahoo.com

Received : 11-May-2021 Accepted : 11-May-2021 Published : 31-May-2021

DOI 10.25259/JHAS_8_2021

Quick Response Code:

Coronavirus-2019, also called Severe Acute Respiratory Syndrome Coronavirus-2019 or SARS-CoV-2 was first reported from China at the end of December 2019 through transmission into man from bat and it produced severe type of pneumonia in the infected people. Within the next month (January 2020), the virus started its world-wide journey after it successfully established the transmission path from man to man and thus created pandemic and caused epidemic. Facing a deadly challenge of the virus, the scientists all over the world, starting from almost zero-level knowledge about the virus, worked hard to know most of characters related to its biology and pathology at molecular level thereby enriching knowledge which helped in development of various tools and technologies to control the virus and develop protection and prevention methods including production of vaccine against the virus. Nevertheless, to exert a better control over the virus, it is necessary to have knowledge of various details about how the virus has been evolved. During the last one year, research work done by the scientists all over the world have produced voluminous data in this area, though in a scattered way. That information indicates that the virus is actively evolving continuously to generate new strains through gain of function mutations for its survival.

knowledge about the mechanisms of evolution of SARS-CoV-2 at molecular level. **Keywords:** SARS-CoV-2, A deadly challenge, Mutant strains, Molecular biology

INTRODUCTION

The Coronoavirus-2019, also called Severe Acute Respiratory Syndrome Coronavirus-2019 or SARS-CoV-2019 was first reported in Wuhan, China and it transmitted into man from bat/ pangolin (Zoonotic origin) and caused a very bad type of pneumonia.[1-3] This virus then started its long widespread journey in the middle of January 2020 and soon reached almost all the countries of the world through transmission from man to man and thus created a pandemic as well as epidemic.^[4] This virus as a single species appears to cause the largest number of human fatalities during the last one year with giving birth to several new strains with gain of certain important functions relating to the infectivity as well as other growth characters as discussed later.

In this short review, I have made an attempt to put together that information to highlight the present status of our

Origin

SARS-CoV-2 is the seventh member in the human coronavirus group that could infect human respiratory organs. The other six are - Coronavirus-229E (HCoV-229E) which appeared in 1965, Coronavirus-OC43 (HCoV-OC43) in 1967, Severe Acute Respiratory Syndrome Coronavirus (SARS-COV) in 2002, Coronavirus NL63 (HCOV-NL-63) in 2004, Coronavirus HKU1 (HKU1) in 2005 and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in 2012. Of these viruses, HCoV-229E and HCoV-OC43 were first isolated from human patients suffering

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms. ©2021 Published by Scientific Scholar on behalf of Journal of Hematology and Allied Sciences

from common cold, while the other five first transmitted from animals like bat, civet or camel. $[4,5]$ Among these seven Coronavirus, HCoV-229E, Coronavirus-OC43, HCOV-NL-63, and HKU1 could produce mild symptoms of common cold while other three SARS-CoV, MERS-CoV, and SARS-CoV-2 can also infect the lower respiratory tract organ producing the symptoms of pneumonia. Again, the SARS-CoV-2 is far more dangerous to humans and produces worse type of pneumonia that may even cause death to the infected person.[1,2] *From detailed studies of the genetic and serological characters, the coronaviruses in* the sub-family *Orthocoronavirida*e *w*ithin the *Coronaviridae* family, have been classified into four different genera namely, *Alphacoronavirus, Betacoronavirus, Gammacoronavirus,* and *Deltacoronavirus*. Coronavirus HCoV-OC43, HCoV-HKU1, SARS-CoV, and SARS-CoV-2 belong to *Betacoronavirus* genus while HCoV-229E and HCoV-NL-63, to *Alphacoronavirus*. [4,6]

Composition of SARS-CoV-2 particles and replication

Each matured particle of SARS-COV-2 contains one molecule single stranded +RNA (ssRNA) containing around 29,900 nucleotide bases wrapped in a nucleocapsid structure formed by the protein N. This nucleocapsid in turn is enclosed in an envelope formed by the assembly of E and M proteins in a lipid bilayer thereby forming a basic spherical structure from the surface of which many spikes are projected outwards giving the particle a crown-like structure, hence its name Coronavirus (Corona = crown).^[5,6] All the seven human Coronaviruses use this spike as their infective organ.[5,7] However, three of them, SARS-COV, HCOV-NL-63, and SARS-COV-2 have evolved their spike protein such that the virus recognize the cell surface protein Angiotensin Converting Enzyme-2 (ACE-2) as their receptor, four others use different proteins as their receptors. The S protein monomer has two distinct functional domains S1 and S2, the former being used to interact with the ACE-2 protein at specific site and the S2, to interact with membrane fused part of ACE-2.^[8-11] Furthermore, during infection of SARS-COV-2, the ACE-2 bound spike protein S is cleaved at around the junction of the S1 and S2 domains by furin like protease.^[12] Almost concomitant with such cleavage event, there occurs subtle changes in the conformations of both ACE-2 bound S2 domain of the virus spike protein and the cellular membrane which ultimately pushes the virus particle inside the cell where uncoating of the particle occurs to release the RNA genome to be used for replication and transcription and other associated biochemical events leading ultimately to form many infective virus particles.^[5]

The 29,900 base-long positive ss RNA molecule codes for about 29 proteins.[13] After entry into the host cell, the RNA molecule is used as mRNA and from its 5' end, two-

third length (about 20,000 bases) synthesized about 16 different proteins (all non-structural) that includes RNA dependent RNA polymerase, transcriptase, several nonstructural proteins RNA helicase, 3'–5' exonuclease, endonuclease, various types of proteases and metal binding proteins and other associated proteins having different functions that are needed for the replication of the genomic RNA molecule.^[5,6] One unique feature of gene expression from the genomic RNA of the coronaviruses is that the 5'-proximal two-third length of the genomic RNA is translated to give two polyproteins 1a and 1b corresponding to ORF1a and ORF1b. During the ongoing translation of 1a, the translating ribosomes makes a negative shift through one base and form a new set of codons at a definite position of mRNA leading to the synthesis of 1b. Both 1a and 1b are polyproteins which are then cleaved by specific proteases at specific sites to generate around 16 proteins having distinct functions as mentioned above.^[5,6] The $3^{\prime}-5^{\prime}$ exonuclease has weak proof reading activity which is used to rectify any misincorporation of nucleotide during RNA replication. The genes coded by the residual one-third RNA at the 3' end proximal segment are transcribed by certain unique mechanisms through the formation of nested mRNAs which are then translated into different proteins that include four structural proteins N, M, E, and S. When the genome RNA is replicated to a large copies and the structural protein are made in sufficient quantities, the matured virus particles are formed by budding and assembly at the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). During such assembly, the virus particles collect their membrane envelope from ERGIC and thus incorporate lipid into their envelope.[5,6] All these virus particles then come out and infect the neighboring cells.

During the first phase of infection, the virus replicates within the cells of nasal cavity and throat and then after about 10 days the virus passes into the lungs and infects pneumocytes and immunocytes. At this stage certain components of the inner immune system of the host are activated to kill the virus. These are: (a) The interferon which exists from earlier infections is used to kill the virus; (b) the amounts of monocytes and macrophages (the two killer cells) and the interleukin molecules (the killer molecules) increases. These killer cells and killer molecules rush to the site of infection and kill the free virus and the infected cells, respectively.^[14]

Evolution of SARS-CoV-2: Gain of function changes in the 'S' gene

Before the appearance of SARS-COV-2, there were six other Coronaviruses which could infect respiratory system including lungs. All these seven viruses use their spike protein 'S' as their infection organ. While keeping this infection

organ same, they have gradually increased both the infection capability and the degree of injury to the host cell during the passage of time from member to member appearing in order. Of these seven viruses only three, SARS-COV, HCOV NL-63, and SARS-COV-2 have evolved to selectively use ACE-2 protein of the host cell as receptor during infection.^[5,7] Again, of these three, the SARS-COV is the first coronavirus which has developed this infection pathway. While HCOV-NL-63 is not so much injurious to host as the other two like SARS-COV and SARS-COV-2, and it is also the fact that SARS-COV-2 is far more effective in producing damage to the host cell after infection.

It has been observed that the affinity with which the spike protein S binds with the ACE-2 protein of the host cell is more than ten-fold higher for SARS-COV-2 protein than the SARS-COV protein. Furthermore, the antibody against the S protein of SARS-COV does not recognize the S protein of SARS-COV-2.[15] These suggest that the SARS-COV-2 has gained more capability of infection and of producing injury through the mutational changes in the S protein.

Extensive comparative analyses of the nucleotide sequence data for SARS-CoV-2 and other related coronaviruses indicate that there occurs recombination events across its genome and that the highest frequency of those events are seen within two specific genes, one in the ORF1a and the other around the N terminus of the gene encoding spike protein S. But those data could not settle between which of those partners were parent and which were recombinant progenies.^[16]

It is a fact that the mutational changes in the S protein within/around the receptor binding domain (RBD) are very important in respect of gain of function of the protein towards infectivity as is seen with SARS-CoV-2. After the first emergence of SARS-CoV-2 in December 2019, during the one year that follows, many new strains of the virus with varying characters have appeared among its progeny and subprogeny population. In all these new strains, the mutational changes have occurred within/around the RBD region of the S protein. These information suggest that the RBD region of S protein is the main target for the evolution of new strains that could maintain their dominance in infectivity and host damage. Based on these information, it may be extrapolated backwards to infer that during the evolution of SARS-CoV-2, various genetic events like recombination between different strains and mutations involving the region encoding S gene might have occurred at different times within bat and the one with gain of function related to increased infectivity and other pathological properties might have been purified by selection during continued transmission in the natural path before jumping into human in December 2019. Several scientists do advocate in favor of the evolution of SARS-CoV-2, through the above processes of recombination and mutation followed by selection in natural ways.^[17-21] The SARS-CoV-2 variants that emerged during last year and their single nucleotide polymorphism (SNP) characters and various properties associated with those mutations are shown in **Table 1**.

Comparative analysis of the sequences of genomic RNAs of SARS-CoV-2 and related viruses predicts the possible evolutionary relationship

*Comp*arison of nucleotide sequences of the genomic RNAs of five samples of SARS-COV-2 isolated from the first five different infected persons in China shows that two of them have sequence similarity to the extent of 99.99% between them, while remaining three have changes at two different positions. On the other hand, the nucleotide sequence of SARS-COV-2 shows 79.9% similarity with that of SARS-CoV and only 51.8% with MERS-COV sequence. When the amino acid sequences deduced from the nucleotide sequences of three different proteins, the envelope protein E, nucleocapsid protein N, and the spike protein S of SARS-COV and SARS-COV-2 were compared, the former two proteins of SARS-COV and SARS-COV-2 show, respectively, 96% and 89.6% similarity while the S protein of SARS-COV-2 shows 77% similarity with that of SARS-COV and only 31.9% with that of the S protein of MERS-COV. So SARS-COV-2 is more close to SARS-COV than to MERS-COV in the evolutionary path. Again two other coronaviruses SL-CovZC45 and SL-CoVZXC21 discovered between 2015 and 2017, both of horse shoe bat origin, have RNA sequence similarity with that of SARS-COV-2 to the extent of 87% and 87.7% while the SARS-COV RNA shows 81% and 81%respectively.[28] So from these information, the SARS-COV-2 may be considered to be evolved from any one of the above two horse shoe bat viruses.

Molecular analyses of the nucleotide sequences of S gene and the amino acid sequence of S protein of SARS-CoV-2 indicate certain unusual features.

Recently it has been shown through further critical analysis of SARS-CoV-2 genome and comparing with those of evolutionarily close viruses like SARS-CoV, RaTG13, and SLCoVZX45 and SLCoVZcX21, that within the segment of RNA carrying the RBD region of S gene, there are certain unusual features in the SARS-CoV-2 genome that are absent in the other above mentioned viral genomes.^[29] These features are: (a) During evolution of SARS-COV-2, its S protein has increased the affinity towards binding to the ACE-2 protein of the host which must be dependent on certain mutational change(s) in its S gene, especially in the segment corresponding to its RBD domain. (b) At the junction of S1 and S2 domains of the S protein there are extra 12 nucleotides (equivalent to 4 codons) by which a polybasic amino acid sequence PRRA (Pro-Arg-Arg-Ala) has been generated at that site where the furin protease can cleave leading to physically separating the S1 and S2 domains during infection. This cleavage event in the S protein is very important for the increase of infection efficiency as well as

the host range of the virus. Gain of this property is also very important for the increase in the infectivity of SARS-COV-2. Besides, the placement of Pro residue at one end of the above furin sensitive site (PRRA) can destabilize the normal secondary structure around the RBD domain of the S protein. (c) The third feature is that in the life history of SARS-CoV-2, the particular region of the S protein that inhabits RBD has suffered relatively more mutations compared to its other two neighboring genes E and N. (d) In the RBD segment of S protein of SARS-COV, the positions of 6 amino acids have been shown to play an important role in the receptor binding function. These are: Y442, L472, N479, D482, T487, and Y491. The amino acids in the RBD segment of the S protein of SARS-COV-2 at the corresponding positions have been shown to be: L455, F486, Q493, S494, N501 and Y505. Therefore, among the 6 amino acids within the receptor binding domain of S protein of SARS-CoV-2, five have been changed compared to those of the S protein of SARS-CoV. (e) The segment of nucleotide sequence of SARS-CoV-2 genome (RNA) that carries the RBD region has been found to be flanked by two restriction enzyme sites, one *Eco*RI site at the 5' side and the other *Bst*EII at the 3' side.

CONCLUSION

In conclusion, the evolution of SARS-COV-2 appears to be caused by so many changes within the short region of viral RNA genome coding for amino acid sequences corresponding to the RBD of the S protein by which it acquired the increased infectivity and the more damaging action to the host cell. But, incorporation of so many mutational changes within the short segment containing the RBD region within S protein raises the question that how by natural process of recombination and mutation all those changes could be possible in the path of evolution of SARS-CoV-2 from whichever ancestor coronavirus it may arise. Further studies are needed to settle this important issue.

ACKNOWLEDGEMENTS

I thankfully acknowledge the help extended by Dr. Nandan Kumar Jana, Assistant Professor, Department of Biotechnology, Heritage Institute of Technology, Kolkata who provided various current references and also helped in verifying certain published data on SARS-CoV-2 genome RNA sequence analysis.

Declaration of patient consent

Patient's consent not required as there are no patients in this study.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- 1. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;579:270–3 DOI:10.1038/ s41586-020-2012-7.
- 2. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, *et al.* A new coronavirus associated with human respiratory disease in China. Nature 2020:579;265–9. https://doi.org/ 10.1038/s41586-020-2008-3.
- 3. Lam TT, Jia N, Zhang YW, Shum MH, Jiang JF, Zhu HC, *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature 2020;583:282–5. https://doi.org/10.1038/ s41586-020-2169-0.
- 4. Singh D, Soojin VY. On the origin and evolution of SARS-CoV-2. Exp Mol Med 2021:53;537–47.
- 5. Artika IM, Dewantari AK, Wiyatno A. Molecular biology of coronaviruses: current knowledge. Heliyon 2020:6;e04743.
- 6. Kadam SB, Sukhramani GS, Bishnoi P, Pable AA, Barvkar VT. SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights. J Basic Microbiol 2021;61:180–202. DOI:10.1002/jobm.202000537.
- 7. Hofmann H, Pyrc K, Van Der Hoek L, Geier M, Berkhout B, Pöhlmann S. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. Proc Natl Acad Sci USA 2005;102:7988–93.
- 8. Li X, Zai J, Zhao Q, Nie Q, Li Y, Foley BT, Chaillon A. Evolutionary history, potential intermediate animal host, and cross‐species analyses of SARS‐CoV‐2. J Med Virol 2020;92:602–11.
- 9. Wang H, Li X, Li T, Zhang S, Wang L, Wu X, *et al.* The genetic sequence, origin, and diagnosis of SARS-CoV-2. Eur J Clin Microbiol Infect Dis 2020;39:1629–35. doi:10.1007/s10096- 020-03899-4.
- 10. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 2020;181:281–92.
- 11. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, *et al*. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 2020;367:1260–3.
- 12. 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. https://doi.org/10.1016/j.antiviral.2020.104742.
- 13. Yoshimoto FK. The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. The Protein J 2020;39:198–216.
- 14. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020;109:102433 doi:10.1016/j. jaut.2020.102433.
- 15. Cao W, Dong C, Kim S, Hou D, Tai W, Du L, *et al.* Biomechanical characterization of SARS-CoV-2 spike RBD and human ACE2 protein-protein interaction. Biophys J 2021;120:1011–9. doi:10.1016/j.bpj.2021.02.007.
- 16. Boni MF, Lemey P, Jiang X, Lam TT, Perry BW, Castoe TA, *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nature Microbiol 2020;5:1408–17. doi:10.1038/s41564-020-0771-4.
- 17. Anderson KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. Proximal origin of SARS-CoV-2 Nat Med 2020;26:450–52. doi:10.1038/s41591-020-0820-9.
- 18. Leitner T, Kumar S. Where did SARS-CoV-2 come from? Mol Biol Evol 2020;37:2463–4.
- 19. Xia X. Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense. Mol Biol Evol 2020;37: 2699–705.
- 20. Zhang YZ, Holmes EC. A genomic perspective on the origin and emergence of SARS-CoV-2. Cell 2020;181:223–7.
- 21. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, *et al*. On the origin and continuing evolution of SARS-CoV-2. Nat Sci Rev 2020;7:1012–23.
- 22. Biswas NK, Majumder PP. Analysis of RNA sequences of 3636 SARS-CoV-2 collected from 55 countries reveals selective sweep of one virus type. Indian J Med Res 2020;151:450–8. DOI:10.4103/ijmr.IJMR_1125_20.
- 23. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, *et al*. Spike mutation D614G alters SARS-CoV-2 fitness. Nature 2021;592:116–21.
- 24. Hou YJ, Chiba S, Halfmann P, Ehre C, Kuroda M, Dinnon KH, *et al*. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science 2020;370:1464–8.
- 25. Lauring AS, Hodcroft EB. Genetic variants of SARS-CoV-2— What do they mean? JAMA 2021;325:529–31. doi:10.1001/ jama.2020.27124.
- 26. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday J, *et al.* Estimated transmissibility and severity of novel SARS-CoV-2 variant of concern 202012/01 in England. medRxiv 2020.12.24.20248822; doi:https://doi.org/ 10.1101/2020.12.24.20248822.
- 27. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y *et al.* Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature 2021;593**:**130–35.
- 28. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, *et al*. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect. 2020;9:221–236. doi:10.1080/22221751.2020.1719902. Erratum in: Emerg Microbes Infect 2020;9:540.
- 29. Yan LM, Kang S, Guan J, Hu S. Unusual features of the SARS-CoV-2 genome suggesting sophisticated laboratory modification rather than natural evolution and delineation of Its probable synthetic route. Zenodo 2020. http://doi. org/10.5281/zenodo.4028830.

DR. NITAI CHANDRA MANDAL did his PhD in Biochemistry from Calcutta University and Post Doctoral Research in the U.S.A in the Department of Molecular Biology, Albert Einstein College of Medicine, New York followed by Department of Microbiology, USC School of Medicine. He then joined Bose Institute, Kolkata and retired as Professor and Emeritus Scientist in 2003. In between, he spent two years (1987–89) as a Visiting Scientist at NIH, Washington D.C., USA for carrying out advanced research in Bacterial Genetics. He received many prestigious awards. He is the elected fellow of *Learned Societies e.g.-* Indian Academy of Sciences, Indian National Science Academy, National Academy of Sciences, Indian Chemical Society and Society of Biological Chemists (India). He used to teach Molecular Genetics, Molecular Biology and Virology to the postgraduates from the discipline of Biochemistry/Microbiology/Biotechnology of Calcutta University, Vidyasagar University, Jadavpur University, IISER (Kolkata), Bose Institute and West Bengal State University. He also served as reviewer of papers for both National & International scientific journals. At different time points he worked as member of of various National Committees. He has published Original research papers in Peer-Review

high impact factor journals like JBC, Virology, Biochemistry, J. General Virology, PNAS (USA), Genes & Development, Microbiology, Molecular and General Genomics, J. Biochem. & Mo. Biol., Protein Engineering, etc. Among many biographical articles/books written by him, the notable area) Acharya P. C. Ray: Life and Times, b) Acharya Jagadish Chandra Bose: 'A Physicist Changed into a Plant Biophysicist', c) The Twilight Years: Understanding Human Aging and Death at Molecular Level (under preparation). He is engaged in writing popular science in Bengali and in National Science Movement being associated with Vivekananda Vijnan Mission, Kolkata. He extended financial help in the form of donations or otherwise to different societies/organizations/ schools that are involved in health care/benevolent jobs/ children's education that includes- Ramkrishna Mission Seva Pratisthan (Kolkata), The Calcutta Heart Clinic & hospital (Salt Lake), Vidya Vikash Trust for blind children's education (Bengaluru), OFFER (that takes care of street children and orphans to bring them to mainstream society), Kolkata and many others.

How to cite this article: Mandal NC. Coronavirus-2019: Origin and evolution. J Hematol Allied Sci 2021;1(1):1–6.