

Genes of SARS-CoV-2 and emerging variants

Sudip Dhakal^A and Ian Macreadie^{A,B}

^ASchool of Science, RMIT University, Bundoora, Vic. 3083, Australia.

^BTel.: +61 402 564 308; Email: ian.macreadie@rmit.edu.au

Abstract. The pandemic caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is distinctly different from outbreaks caused by other coronaviruses: SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV). The differences in the rapid transmission and severity of human coronaviruses are due to the genetic composition of the virus. SARS-CoV-2 contains genes encoding non-structural proteins (NSPs), structural proteins, and accessory proteins. The NSPs are mainly involved in replication of the virus within the host and inhibition of the host defence system. Structural proteins are involved in viral entry and attachment to host cells, preservation of the core virion and elicit the majority of the immune response. The functions of the accessory proteins are largely unknown. Most focus has been given to structural proteins, especially the spike protein as the strongest vaccine candidate. However, the recent emergence of spike variants and their ability to rapidly transmit and escape neutralisation by vaccine-induced antibodies has threatened the global community. Meanwhile, recent studies of accessory proteins reveal their importance in viral pathogenesis. Hence, proper understanding of the functions of all unknown viral proteins is crucial to devise alternate antiviral strategies.

Received 15 February 2021, accepted 24 February 2021, published online 9 April 2021

Introduction

The COVID-19 pandemic, caused by the member of Coronaviridae family, *Orthocoronavirinae* subfamily, *Betacoronavirus* genus and subgenus *Sarbecovirus*, has severely affected human lifestyles all around the globe¹. Owing to its high similarity to the severe acute respiratory syndrome coronavirus (SARS-CoV), the virus was named SARS-CoV-2. Disease caused by the virus varies from asymptomatic to severe and frequently fatal pneumonia. Inadequate measures to control the virus, combined with its high infectivity, have had serious consequences for public health worldwide.

Genes of SARS-CoV-2

Coronaviruses are small, 80-220 nm diameter, enveloped, positive sense, single-stranded RNA viruses that replicate within the cytoplasm of host cells using the host endoplasmic reticulum and Golgi network for processing of the proteins and assembly of the virus^{2,3}. The 30 kb genome sequence of the SARS-CoV-2 virus has 15 open reading frames (ORF) that encode nearly 29 proteins (Figure 1)^{2,4}. At the 5' end of the viral genome are ORF1a and ORF1b, encoding polyproteins that are proteolytically processed into 16 non-structural proteins (NSP1 to NSP16). The non-structural proteins play vital roles in creating a suitable environment for viral invasion and replication by hijacking the host protein synthesis

machinery, inhibiting host mRNA expression and antiviral defences⁵. Detailed descriptions of the functions of the majority of the SARS-CoV-2 proteins are reviewed elsewhere⁴. At the 3' end are 13 ORFs that encode four structural proteins (spike protein (S), envelope protein (E), nucleocapsid protein (N) and membrane protein (M)) and nine accessory proteins. Among the structural proteins, the trimeric S protein (composed of two subunits S1 and S2) is crucial for allowing viral attachment. It interacts with the host cell angiotensin converting enzyme 2 (ACE2) receptor via a receptor binding domain (RBD) present in the S1 subunit allowing the viral attachment to host cells and aiding in the fusion of the membrane via S2 subunit³. Importantly the S protein is also the major viral protein eliciting both humoral and cell-mediated immune responses, making it an excellent vaccine candidate⁶. However, mutations in the S protein pose a significant threat by possible generation of mutant strains with the potential to escape antibody neutralisation, thus rendering the current vaccine strategy less effective⁷⁻¹⁰.

Emerging variants of concern

Several lineages of variants have emerged since the first case in Wuhan, China in December 2019, posing greater risks to the global community due to specific mutations in their spike proteins, causing the strains to become more virulent and resistant to neutralising antibodies¹¹. The most striking ones are those with mutations at the



Figure 1. Schematic diagram showing genome of SARS-CoV-2. Green coloured genes encode polyproteins that are proteolytically cleaved into non-structural proteins (NSP), red coloured genes encoding structural proteins and blue coloured genes encoding accessory proteins.

Table 1. Emerging SARS-CoV-2 variants of concern and their effect on viral pathogenesis.

Emerging variants	Lineage	Spike protein mutation	Viral pathogenesis of variant	Reference
South African variant	20H/501Y.V2 or B.1.351	L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V	Increased infectivity and escape from human neutralising antibodies	9,10
		Potential deletion at L242_L244 or mutation L242H		
Italian variant	MB61	K182N, Q493K, N501T, D614G	Alteration in ACE2 binding and escape from vaccine induced or natural neutralising antibodies	12
		L242 deletion		
UK variant	20I/501Y.V1, VOC 202012/01, or B.1.1.7	N501Y, A570D, P681H, T716I, S982A, D1118H	Increased transmissibility, disease severity	8,13
		Y144 deletion, H69_V70 deletion		
Brazilian variant	P.1 or 20J/501Y.V3	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, T1027I	Alteration in ACE2 interaction and human antibody; increased resistance to neutralising antibodies	8,14

receptor binding domains, which not only alter the interaction of the spike proteins with the ACE2 receptor but also reduce affinity towards human neutralising antibodies (refer to Table 1). Examples of such variants are the UK variant (B.1.1.7), South African variant (B.1.351) and the Brazilian variant (P.1), that are currently emerging as increased threats worldwide.

Future directions

Among the various mutations in new variants of SARS-CoV-2, the spike mutations K417N, E484K, N501Y, N501T and D614G appear crucial for determining critical alterations in viral behaviour. Mutations that occur in the RBDs of spike proteins cause serious concerns as current vaccination strategies are based on the original spike protein. Currently, there are no antiviral agents that can effectively inhibit SARS-CoV-2 replication in humans¹⁵. Considering these limitations, it is of utmost importance to explore new ways towards finding strategies to reduce viral proliferation. Some NSPs, non-spike structural proteins and accessory proteins could be investigated as these proteins are involved in important steps of viral replication and pathogenesis such as viral protein synthesis, hijacking host protein synthesis machinery and inhibiting intracellular host defence mechanisms⁴. Some such protein targets are NSP1 (inhibits host mRNA translation), NSP3 (papain like protease that releases

NSP1, NSP2 and NSP3 from viral polyprotein), NSP5 (main protease that cleaves viral polyproteins into individual functional proteins), NSP6 (prevents host autophagosome-lysosome fusion) and N-protein (preserves viral RNA from interfering RNAs from host).

The functions of the accessory proteins of SARS-CoV-2 are largely unknown and further investigations uncovering their role may shed light on the unanswered questions of rapid spread and differences in disease severity of SARS-CoV-2. Several accessory proteins are being investigated for their role during viral invasion. In a recent study comparison of the proteins encoded by SARS-CoV-2 with the proteins of SARS-CoV revealed the presence of novel ORF8 and ORF10 proteins in SARS-CoV-2. Structural analysis of ORF8 protein supported the claim that the protein can form unique large-scale assemblies that are not present in SARS-CoV, potentially inhibiting the host defence system¹⁶. Additionally, association of ORF8 deletion variant (Δ 382 variant) with milder disease outcome strongly supports the importance of ORF8 protein as a therapeutic target against SARS-CoV-2¹⁷. Meanwhile, ORF10 is the only protein that is not present in other human coronaviruses and is yet to be studied. Nucleotide sequences similar to ORF10 were found in SARS-CoV but with a nonsense mutation, causing early termination of protein synthesis¹⁸. Further analysis of the predicted structure of the ORF10 protein showed a signal peptide and a hydrophobic

β -sheet region beyond the cleavage site, suggesting a possible hydrophobic interaction and membrane localisation, maybe as transport protein. These findings suggest ORF10 protein's possible role in viral invasion, which needs to be explored further¹⁸.

Other accessory proteins differ in their effect and activity despite a high similarity in protein sequence alignment. A recent study comparing the ability of ORF3a proteins of SARS-CoV-2 and SARS-CoV to induce apoptosis showed reduced induction of cell death by SARS-CoV-2 protein¹⁹. These differences in the viral proteins could explain the differences in the severity of the SARS-CoV-2 and SARS-CoV. Further research and new insights in the potential role of accessory proteins during viral invasion could indicate novel ways to reduce the infectivity, rapid spread and boost host immune system. Such studies will not only lead to therapeutic strategies against SARS-CoV-2 virus, but also be useful for similar viral infections including SARS-CoV and MERS-CoV infections.

Conflicts of interest

Ian Macreadie is the Editor-in-Chief of *Microbiology Australia* but was blinded from the peer-review process for this paper.

Acknowledgements

This research did not receive any specific funding.

References

- Mackenzie, J.S. and Smith, D.W. (2020) COVID-19: a novel zoonotic disease caused by a coronavirus from China: what we know and what we don't. *Microbiol. Aust.* **41**, 45–50. doi:10.1071/MA20013
- V'kovski, P. *et al.* (2021) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170. doi:10.1038/s41579-020-00468-6
- Howard-Jones, A.R. and Kok, J. (2020) The SARS-CoV-2 'perfect storm': from humble betacoronavirus to global pandemic. *Microbiol. Aust.* **41**, 150–156. doi:10.1071/MA20040
- Yoshimoto, F.K. (2020) The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J.* **39**, 198–216. doi:10.1007/s10930-020-09901-4
- Zinzula, L. (2021) Lost in deletion: the enigmatic ORF8 protein of SARS-CoV-2. *Biochem. Biophys. Res. Commun.* **538**, 116–124. doi:10.1016/j.bbrc.2020.10.045
- Yang, J. *et al.* (2020) A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature* **586**, 572–577. doi:10.1038/s41586-020-2599-8
- Thomson, E.C. *et al.* (2021) Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **184**, 1171–1187.e20. doi:10.1016/j.cell.2021.01.037
- Fratev, F. (2020) The N501Y and K417N mutations in the spike protein of SARS-CoV-2 alter the interactions with both hACE2 and human derived antibody: a free energy of perturbation study. *bioRxiv* [Preprint]. doi:10.1101/2020.12.23.424283
- Tegally, H. *et al.* (2020) Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* [Preprint]. doi:10.1101/2020.12.21.20248640
- Jangra, S. *et al.* (2021) The E484K mutation in the SARS-CoV-2 spike protein reduces but does not abolish neutralizing activity of human convalescent and post-vaccination sera. *medRxiv* [Preprint]. doi:10.1101/2021.01.26.21250543
- Burki, T. (2021) Understanding variants of SARS-CoV-2. *Lancet* **397**, 462. doi:10.1016/S0140-6736(21)00298-1
- Fiorentini, S. *et al.* (2021) First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August, 2020. *The Lancet Infectious Diseases*. doi:10.1016/S1473-3099(21)00007-4
- Zhang, L. *et al.* (2020) The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* [Preprint]. doi:10.1101/2020.06.12.148726
- Wang, P. *et al.* (2021) Increased resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7 to antibody neutralization. *bioRxiv* [Preprint]. doi:10.1101/2021.01.25.428137
- Singh, K.P. *et al.* (2020) Therapeutics for COVID-19: established and in development. *Microbiol. Aust.* **41**, 217–223. doi:10.1071/MA20058
- Flower, T.G. *et al.* (2021) Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. *Proc. Natl. Acad. Sci. USA* **118**, e2021785118. doi:10.1073/pnas.2021785118
- Young, B.E. *et al.* (2020) Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* **396**, 603–611. doi:10.1016/S0140-6736(20)31757-8
- Dhokal, S. *et al.* (2020) Could the severity of COVID-19 be enhanced by ORF10 accessory proteins? *Curr. Top. Pept. Protein Res.* **21**, 97–106.
- Ren, Y. *et al.* (2020) The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cell. Mol. Immunol.* **17**, 881–883. doi:10.1038/s41423-020-0485-9

Biographies



Sudip Dhakal is a PhD candidate at School of Science, RMIT University under Biosciences discipline, whose research focuses on overcoming reduced proteostasis in Alzheimer's disease.



Ian Macreadie is an Honorary Professor at RMIT University and Editor-in-Chief of *Microbiology Australia*. His research and expertise are in diverse fields of biosciences ranging from industrial microbiology to biomedical research.

Member of Australian Society for Microbiology (MASM)

Want to upgrade from an associate to a full member?

No points or cash required and only one supporting referee needed.

Go to the ASM website, download and complete the form, and forward it with a CV to the National Office.

We will do the rest.