



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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Forum

Molecular Determinants of SARS-CoV-2 Variants

Arinjay Banerjee,<sup>1,2,3,\*</sup>  
Karen Mossman,<sup>4,5,6</sup> and  
Nathalie Grandvaux<sup>7,8,\*</sup>



**Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) evolution is expected, given the nature of virus replication. Selection and establishment of variants in the human population depend on viral fitness and on molecular and immunological selection pressures. Here we discuss how mechanisms of replication and recombination may contribute to the emergence of current and future variants of SARS-CoV-2.**

SARS-CoV-2 emerged in December 2019 to cause a worldwide pandemic. Rapid sharing of information by researchers enabled the development of vaccines in record time. However, as SARS-CoV-2 continues to spread, it is evolving. This evolution should come as no surprise since RNA viruses are known to mutate, given their error-prone replication process. As SARS-CoV-2 circulates, mounting selection pressures, such as natural infection- and vaccine-mediated immunity will influence the selection of viral variants. In this forum article we discuss how two processes – replication and recombination – contribute to the generation of current and future SARS-CoV-2 variants.

**Replication-Induced Genomic Errors**

Coronaviruses (CoVs), such as SARS-CoV-2, are RNA viruses that use an RNA-dependent RNA polymerase (RdRp) to generate genomic copies. The RdRp is intrinsically error-prone, thus promoting

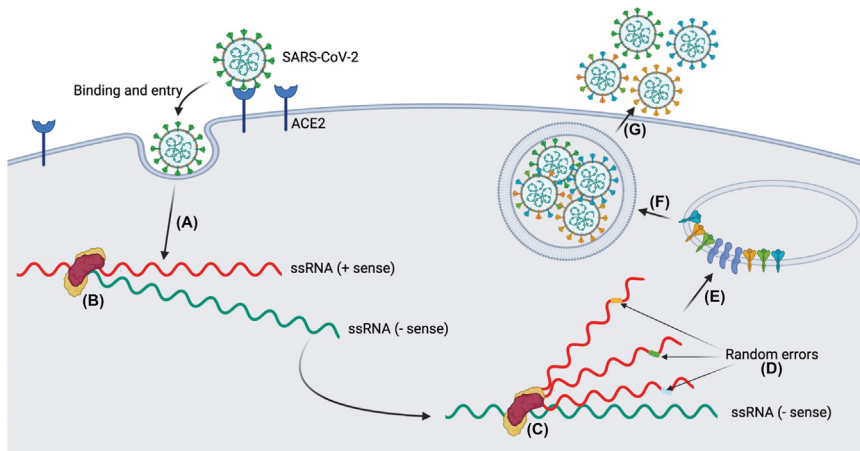
virus evolution through replication-associated changes (Figure 1). The mutation rate is kept low by the proofreading activity of a virus-encoded 3' exonuclease, such as nsp14 for SARS-CoV-2. Despite proofreading activity, SARS-CoV-2 accumulates genetic changes over time as evidenced by emerging variants. Understanding the molecular factors that lead to the generation of SARS-CoV-2 variants, along with extrinsic factors that lead to the selection and establishment of these variants, is critical to control the ongoing pandemic. The coronavirus disease 2019 (COVID-19) pandemic has also reignited interest in studying wildlife reservoirs. Selection analyses suggest that most adaptive changes within SARS-CoV-2 genomes occurred in wildlife reservoirs, likely bats, prior to emerging in human populations [1]. It would be highly valuable to assess the rate of accumulation of replication-induced errors in the genomes of SARS-CoV-2 or SARS-CoV-2-related viruses in other mammalian species, such as bats, to determine the true diversity of these viruses in wildlife reservoirs. It remains unknown if infected bats produce a more diverse quasi-variant population of sarbecoviruses compared with other permissive mammalian species.

The biochemical make-up of different cell types from different hosts may impact CoV replication and rate of error. For example, the catalytic reaction of viral polymerases requires certain substrates, the quality and quantity of which may differ between cell types or cells from alternative hosts [2]. Host factors, including enzymes, can also impact the evolution of viral variants. Inherent differences in the coding sequence of host enzymes, such as TMPRSS2, become more prominent between distantly related mammalian species. When a zoonotic virus such as SARS-CoV-2 spills over into a new mammalian species, the virus interacts with a different ortholog of TMPRSS2. Although the spike protein may be

sufficiently cleaved in this new host to facilitate cellular entry, over time the virus population is likely to evolve and select for variants with higher cleavage efficiency with the TMPRSS2 ortholog in the new animal species [3]. Indeed, furin cleavage sites in SARS-CoV-2-related CoVs (SARS2r-CoVs), discovered in bats, differ in their sequence from human isolates of SARS-CoV-2 [4], suggesting possible species-specific adaptation in host protease utilization by the virus.

Accumulation of genomes with replication-associated random mutations within the same cell has the potential to generate a heterogeneous mix of viral proteins, including spike protein, which brings into question the true diversity of the spike on the surface of individual virions and subsequent tissue-level spread of SARS-CoV-2 and its variants. If a more orderly molecular process to regulate the homogeneous distribution of spike protein variants on individual virions exists, it remains to be identified. Regular RNA sequencing analyses are unable to identify these differences since they can only be captured by ultra-high-resolution single-virion protein sequencing, the technology for which is in its infancy.

The SARS-CoV-2 genomic mutation rate in humans is estimated at  $0.8\text{--}2.38 \times 10^{-3}$  nucleotide substitutions per site per year largely based on analysis of sequencing data archived in public repositories [5]. Emerging studies are attempting to confirm this mutation rate using experimental investigations [6]. By comparison, mutations rates for influenza A virus and Middle East respiratory syndrome CoV (MERS-CoV) are  $2.3 \times 10^{-3}$  and  $1.12 \times 10^{-3}$  nucleotide substitutions per site per year, respectively [7–10]. Human CoVs (HCoV)-OC43 and HCoV-229E have an average mutation rate of  $3\text{--}6 \times 10^{-4}$  nucleotide substitutions per site per year [8]. Emerging experimental data suggest that SARS-CoV-2 is capable of mutating and accumulating changes when facing a new cell type, albeit in the



Trends in Microbiology

**Figure 1. Replication-Associated Generation of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variants.** Replication-associated mutations and intrinsic selection factors play a critical role in the evolution and subsequent emergence of SARS-CoV-2 variants from infected cells. SARS-CoV-2 genome [positive (+) sense single stranded (ss)RNA] is released into the cytoplasm after virus binding to the angiotensin-converting enzyme 2 (ACE2) receptor and entry (A; red strand). The viral replicase complex transcribes the genome into negative (-) sense ssRNA (B; green strand) which acts as a template for subsequent synthesis of full-length +sense ssRNA genome copies (C). During replication, errors introduce mutations within the SARS-CoV-2 genome (D), which lead to mutations within viral proteins (E). While one copy of the genome is packaged into each SARS-CoV-2 virion (F), the heterogeneity of spike proteins on the surface of each virion (represented by different colors) remains unknown. Finally, SARS-CoV-2 variants carrying +sense ssRNA genomes with a varying number of mutations are released from the infected cell following egress (G).

absence of immune surveillance in a single-cell-type infection model [6]. A clinical study also reported the rapid evolution of SARS-CoV-2 variants in the presence of antibodies from convalescent plasma therapy [11]. While random errors during replication may induce genetic mutations in SARS-CoV-2, multiple extrinsic factors, such as individual and population-level immunity, play a vital role in the selection of these variants. More research is warranted to fully understand the cellular and molecular drivers of genomic mutation and selection in SARS-CoV-2. As SARS-CoV-2 accumulates new genetic changes, we shall need to reassess the mutation rate to better understand the contribution of replication-associated random mutations and its impact on SARS-CoV-2 transmission and emergence of new variants.

### Recombination-Associated Mutations

While replication-associated mutations generate small changes in the viral

genome, recombination may introduce more important modifications, leading to dramatic changes in the phenotype of SARS-CoV-2 (Figure 2). Discontinuous transcription of CoV genomes enables recombination in a cell coinfecting with more than one CoV species or variant via ‘strand switching’ by the viral RdRp (Figure 2A) [12]. This process leads to the production of chimeric subgenomic RNA and proteins. Chimeric proteins may have consequences for the fate of infected cells and for cellular and tissue tropism of progeny virions. As technology evolves, we might be able to track the spread of individual virions, along with deciphering the heterogeneity of their composition.

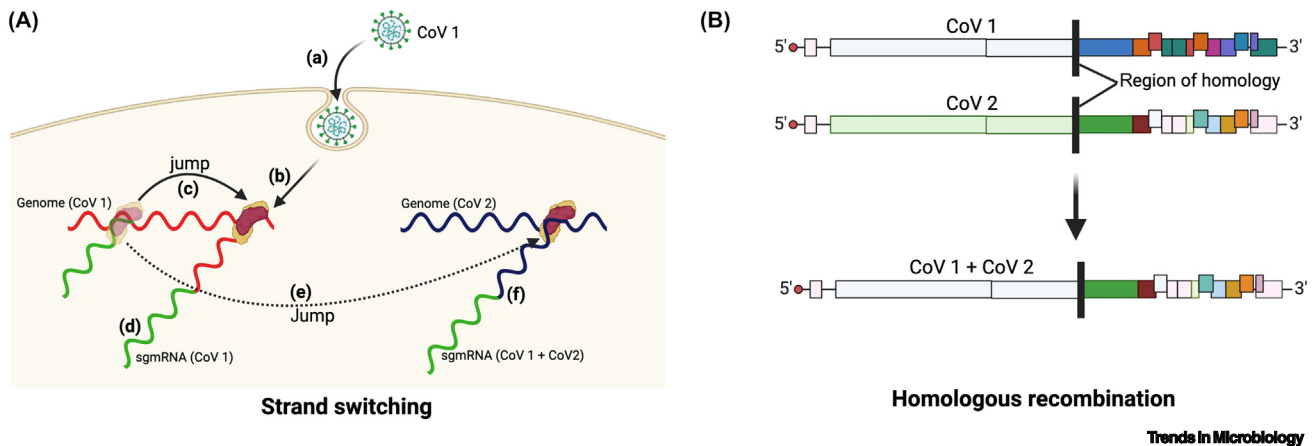
SARS-CoV-2 genomes can also undergo homologous recombination, in which RNA molecules recombine via regions of high similarity (Figure 2B). Recombination events are common for beta-CoVs and have been observed between SARS-related CoVs in wild-caught bats [13,14].

Data suggest that recombination within HCoVs, including SARS-CoV-2, does occur [15]. Recently, eight potential recombination events among SARS-CoV-2 genomes have been identified. Four of eight were noted within structural genes, while two of eight were noted within the spike gene. It is important to note that these recombination events were not found across all subsampled datasets [15], and the frequency of SARS-CoV-2 recombination and factors that influence the rate of recombination still remain elusive. It is vital to monitor SARS-CoV-2 recombination as part of ongoing surveillance and diagnostic efforts, especially among circulating variants and also with other beta-CoVs. These studies will be crucial to assess the evolutionary trajectory of SARS-CoV-2 within the ongoing pandemic, along with predicting the long-term efficacy of vaccine- and natural infection-mediated immunity.

### The Need to Investigate the Molecular Determinants of SARS-CoV-2 Variant Emergence

Tissue and cell tropism, and receptor distribution, are important determinants of coinfection by two closely related CoVs, and hence, the probability of recombination. However, little is known about other viral and host factors that are required to permit or facilitate CoV recombination. Much work remains to be done to fully characterize the replication machinery of SARS-CoV-2 and the fidelity of its RdRp in different species. Understanding the replication process will be critical to determine if SARS-CoV-2 is more prone to replication-induced errors and/or recombination, relative to other highly pathogenic CoVs, such as SARS-CoV and MERS-CoV.

The ongoing COVID-19 pandemic has thrust CoV research to the forefront and we are making tremendous progress in understanding coronavirus pathogenesis in general. While the focus so far has



**Figure 2. Recombination-Induced Variants.** (A) Strand switching. Discontinuous transcription facilitates the generation of potential chimeric subgenomic RNA (sgmRNA) in cells that are infected with more than one species or variant of Coronavirus (CoV). A CoV (CoV 1) entering (a) an already-infected cell will release its genomic RNA in the cytoplasm (b), which will undergo discontinuous transcription (c) to generate sgmRNA (d). During discontinuous transcription, the viral RNA-dependent RNA polymerase (RdRp) may inadvertently switch strands and jump (e) onto a template from a second CoV (CoV 2), leading to the synthesis of chimeric sgmRNA (CoV1+CoV2) (f). (B) Homologous recombination. Similar to recombination between homologous DNA sequences, regions of homology between two CoV genomes (CoV1 and CoV2) may facilitate breakage and rejoining of genomes to form chimeric genomes (CoV1+CoV2).

largely been on developing vaccines and therapeutics against SARS-CoV-2, as effective vaccines are rolling out, researchers are focusing on the biology of SARS-CoV-2. Many questions about SARS-CoV-2 biology still remain – do some host or selected cell types favor recombination over others? Do individual virions carry multimeric or a heterologous set of spike variants, and if so, how does that affect virus spread and multiorgan dispersion? More research is certainly warranted to identify molecular hotspots and catalysts of SARS-CoV-2 replication- or recombination-induced mutants and the pathophysiological consequences of infection with these mutants.

### Acknowledgments

Research on SARS-CoV-2 in the laboratory of N.G. is supported by funds from Fonds de Recherche du Québec – Santé (FRQ-S), Ministère de l'Économie et de l'Innovation – Québec (MEI) and Fondation du CHUM. Research in the laboratory of K.M. is supported by funds from the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council (NSERC) and the Canadian Foundation for Innovation. A.B. was the recipient of a fellowship from NSERC. This work was also supported by CIHR COVID-19 priority funding awarded to principal investigator K.M. and coinvestigator A.B. VIDO

receives operational funding for its CL3 facility (InterVac) from the Canada Foundation for Innovation through the Major Science Initiatives. VIDO also receives operational funding from the Government of Saskatchewan through Innovation Saskatchewan and the Ministry of Agriculture. This article is published with the permission of the Director of VIDO, journal series no. 938. N.G. is a member of the Quebec Respiratory Health Research Network and the Quebec COVID-Pandemic Network.

### Declaration of Interests

There are no interests to declare.

<sup>1</sup>Vaccine and Infectious Disease Organization, University of Saskatchewan, Saskatoon, SK, S7N 5E3, Canada

<sup>2</sup>Department of Veterinary Microbiology, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, SK, S7N 5B4, Canada

<sup>3</sup>Department of Biology, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

<sup>4</sup>Department of Medicine, McMaster University, Hamilton, ON, L8S 4K1, Canada

<sup>5</sup>Michael G. DeGroot Institute for Infectious Disease Research, McMaster University, Hamilton, ON, L8S 4K1, Canada

<sup>6</sup>McMaster Immunology Research Centre, McMaster University, Hamilton, ON, L8S 4K1, Canada

<sup>7</sup>Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montréal, QC, H2X 0A9, Canada

<sup>8</sup>Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montréal, QC, H3C 3J7, Canada

\*Correspondence:

arinjay.banerjee@usask.ca (A. Banerjee) and  
nathalie.grandvaux@umontreal.ca (N. Grandvaux).

<https://doi.org/10.1016/j.tim.2021.07.002>

© 2021 Elsevier Ltd. All rights reserved.

### References

- MacLean, O.A. *et al.* (2021) Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. *PLoS Biol.* 19, e3001115
- Choi, K.H. (2012) Viral polymerases. *Adv. Exp. Med. Biol.* 726, 267–304
- Qing, E. and Gallagher, T. (2020) SARS coronavirus redux. *Trends Immunol.* 41, 271–273
- Wacharapluesadee, S. *et al.* (2021) Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* 12, 972
- De Maio, N. *et al.* (2021) Mutation rates and selection on synonymous mutations in SARS-CoV-2. *Genome Biol. Evol.* 13, evab087
- Borges, V. *et al.* (2021) Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. *bioRxiv* Published online May 19, 2021. <https://doi.org/10.1101/2021.05.19.444774>
- Zhao, Z. *et al.* (2004) Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol. Biol.* 4, 21
- Wang, Y. *et al.* (2018) Current understanding of middle east respiratory syndrome coronavirus infection in human and animal models. *J. Thorac. Dis.* 10, S2260–S2271
- Shu, L.L. *et al.* (1993) Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. *J. Virol.* 67, 2723–2729
- Bar-On, Y.M. *et al.* (2020) SARS-CoV-2 (COVID-19) by the numbers. *eLife* 9, e57309
- Kemp, S.A. *et al.* (2021) SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 592, 277–282
- Lai, M.M. *et al.* (1985) Recombination between nonsegmented RNA genomes of murine coronaviruses. *J. Virol.* 56, 449–456
- Boni, M.F. *et al.* (2020) Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat. Microbiol.* 5, 1408–1417
- Hu, B. *et al.* (2017) Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* 13, e1006698
- Pollett, S. *et al.* (2021) A comparative recombination analysis of human coronaviruses and implications for the SARS-CoV-2 pandemic. *bioRxiv* Published online March 8, 2021. <https://doi.org/10.1101/2021.03.07.434287>