

The SARS-CoV-2 mutation landscape is shaped before replication starts

Authors: Diego Masone^{1,2}, Maria Soledad Alvarez^{3*}, Luis Mariano Polo^{1*}

Affiliations:

¹ Instituto de Histología y Embriología de Mendoza (IHEM) - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Cuyo (UNCuyo), Mendoza, Argentina.

² Facultad de Ingeniería, Universidad Nacional de Cuyo (UNCuyo), Mendoza, Argentina.

³ Instituto de Medicina y Biología Experimental de Cuyo (IMBECU), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Cuyo (UNCuyo), Mendoza, Argentina.

* Corresponding authors. Emails: msalvarez@mendoza-conicet.gob.ar, lpolo@mendoza-conicet.gob.ar

Abstract

Mutation landscapes and signatures have been thoroughly studied in SARS-CoV-2. Here, we analyse those patterns and link their changes to the viral replication niche. Surprisingly, those patterns look to be also modified after vaccination. Hence, by deductive reasoning we identify the steps of the coronavirus infection cycle in which those mutations initiate.

Main

Modifications in the mutation landscape of a genomic sequence can result through several mechanisms¹, such as error-prone polymerases, metabolism, and damaging agents, like an unbalanced redox environment. The comprehensive analysis of the SARS-CoV-2 interhost single base substitution (SBS) showed a mutational spectrum dominated by C>U and, surprisingly, G>U substitutions²⁻⁵. Our results confirmed that pattern (Fig.1A), followed by G>A and A>G, as expected. Transition-type SBS –the interchanges between purines (C>U and U>C) or pyrimidines (G>A and A>G)– were expected to be the most frequent, as they can result from the activity of antiviral enzymes such as APOBEC and ADAR deaminases in host cells⁵⁻⁷. Indeed, uracil is the outcome of cytidine deamination. In contrast, G>U transversions, particularly prevalent in SARS-CoV-2⁸, can result from stochastic processes, such as the misincorporation of nucleotides by an error-prone polymerase with a specific bias or the chemical modification of RNA. Those hypotheses have been discussed previously^{3,4,9,10}, and it is widely agreed that G>U transversion is caused by mutagen exposure, like oxidation due to reactive oxygen species (ROS). That effect was observed not only in viruses but also in bacteria¹¹. Thus, the process begins with the oxidation of a guanine base to produce 8-oxoguanine (8-oxoG). Like guanine, 8-oxoG can pair with cytosine; however, it can also pair with adenine (Fig.1B). Exceptionally, if 8-oxoG pairs with adenine during the first cycle of viral RNA replication, it can be substituted by uracil in the second replication cycle.

Here, we considered how that misincorporation could occur during the intracellular life cycle of SARS-CoV-2. Various external mechanisms can explain modifications in the redox balance in infected cells, with the immune system as the prime suspect¹². The immune cells and those regulating their functions vary through the respiratory tract^{13,14}. Therefore, SBS patterns from those lineages infecting only part of the respiratory tract should differ from those that can infect the whole tract. For example, omicron subvariants (BA.1 and BA.2) mainly replicate in the upper respiratory tract (URT)¹⁵, which reflects in a significant decrease in the G>U/C>A ratio when compared to Alpha and Delta (Fig 1A and Fig1C; and Ruis *et al.*¹⁶). Furthermore, if the immune system were indeed responsible for the changes in the oxidative environment of the infected cells^{12,17}, differences would be expected between SBS spectra from unvaccinated and vaccinated patients^{18,19}. Therefore, we analysed samples from patients infected with alpha and delta variants divided into two groups, unvaccinated and vaccinated. Remarkably, there were significant alterations in the mutation landscape of SARS-CoV-2 genomes from vaccinated patients. Those alterations circumscribe to a rise in G>U transversion (Fig.1D), sustaining the importance of the immunological responses to the oxidative nature of those mutations.

Then, we hypothesised two scenarios where the nucleotide mispairing could occur when vRNA is outside or inside double-membrane vesicles (DMVs), leading to different substitution patterns (Fig.2A and 2B). SARS-CoV-2 contains a positive non-segmented RNA genome [(+)vRNA]. Its replication comprises the early translation of a large polypeptide, then cleaved to produce the RNA-dependent RNA polymerase (RdRp). Both (+)vRNA and RdRp are compartmentalised into DMVs²⁰, avoiding the action of nucleases during vRNA replication²¹. vRNA is then processed through double-stranded RNA intermediates in a sophisticated manner involving (+)vRNA and (-)vRNA. Nevertheless, some vRNA molecules generated inside DMVs are transported to the cytoplasm to produce viral structural proteins. In this scenario, G>U and C>A should have similar magnitudes if the mispairing occurs during replication inside DMVs (Fig.2A). However, G>U substitutions prevail over C>A (Fig.1A, 1C and 1D), favouring the theory where the mispairing happens before the vRNA is enclosed into a DMV (Fig.2B). Subsequently, the asymmetry between G>U and C>A transversions can be explained by inferring that guanine oxidation occurs mainly outside DMVs (Fig.2A), so compartmentalisation can play a role in decreasing the exposure of vRNA to the oxidative environment, protecting it from ROS action. Similarly, other coronaviruses also shield their replication processes and machinery by using DMVs²². Consequently, it is unsurprising that the unbalance between G>U and C>A was previously observed in SARS-CoV and MERS-CoV⁵. Interestingly, the C>U/G>A ratio yields a value comparable to G>U/C>A (Fig.2C), suggesting that enzymatic deamination also happens before getting into the DMVs, and supporting the concept of the protective role of compartmentalisation.

Additional studies are needed to elucidate in detail mechanisms driving viral mutation patterns and how that drive evolution of new SARS-CoV-2 strains. Particularly, if vaccines could cause novel strains appearance or to affect viral fitness through those mutations, further investigations are warranted to uncover how to manipulate that effect favouring their efficacy.

Methods

Calculation of SBS spectra

We calculated SBS mutational spectra for SARS-CoV-2 lineages using the 29th of July 2022 UShER SARS-CoV-2 phylogenetic tree²³, as described in Ruis 2022¹⁶, with the small modifications described below.

To calculate SBS spectra for each vaccination status, we identified mutations on tip phylogenetic branches leading to sequences with known vaccination status in their GISAID metadata.

We calculated confidence intervals for mutation type proportions through the Wilson score interval using the total number of mutations as the number of trials and the proportion of the mutation type as the success proportion.

Data availability

The findings of this study are based on metadata available on GISAID²⁴, via gisaid.org/EPI_SET_220927fu (doi: 10.55876/gis8.220927fu) and [/EPI_SET_220925np](https://gisaid.org/EPI_SET_220925np) (doi: 10.55876/gis8.220925np), for unvaccinated and vaccinated patients (respectively).

Acknowledgements

The authors thank Dr Stuart Rulten, Prof Luis Mayorga and Prof Claudia Tomes for their critical and useful comments while conducting this study, and Dr Christopher Ruis (MRC-LMB, Cambridge, UK) and Dr Angie S. Hinrichs (University of California Santa Cruz, Santa Cruz, CA, USA), for their technical support. We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. Funding: M.S.A. is a postdoctoral fellow of CONICET. This work was supported by grants from ANPCyT PICT2019-01889 and CONICET-PIP3195 to L.M.P.; ANPCyT PICT2020-01897 and CONICET-PIP0409 to D.M.

Competing interests

The authors declare that they have no competing interests.

References

- 1 Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**, 821-836 e816, doi:10.1016/j.cell.2019.03.001 (2019).
- 2 Popa, A. *et al.* Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci Transl Med* **12**, doi:10.1126/scitranslmed.abe2555 (2020).

- 3 Panchin, A. Y. & Panchin, Y. V. Excessive G-U transversions in novel allele variants in SARS-CoV-2 genomes. *PeerJ* **8**, e9648, doi:10.7717/peerj.9648 (2020).
- 4 De Maio, N. *et al.* Mutation Rates and Selection on Synonymous Mutations in SARS-CoV-2. *Genome Biol Evol* **13**, doi:10.1093/gbe/evab087 (2021).
- 5 Di Giorgio, S., Martignano, F., Torcia, M. G., Mattiuz, G. & Conticello, S. G. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv* **6**, eabb5813, doi:10.1126/sciadv.abb5813 (2020).
- 6 Liu, R. *et al.* Genomic epidemiology of SARS-CoV-2 in the UAE reveals novel virus mutation, patterns of co-infection and tissue specific host immune response. *Sci Rep* **11**, 13971, doi:10.1038/s41598-021-92851-3 (2021).
- 7 Li, J. *et al.* Two-step fitness selection for intra-host variations in SARS-CoV-2. *Cell Reports*, In press, doi:10.1016/j.celrep.2021.110205 (2022).
- 8 Forni, D., Cagliani, R., Pontremoli, C., Clerici, M. & Sironi, M. The substitution spectra of coronavirus genomes. *Brief Bioinform*, doi:10.1093/bib/bbab382 (2021).
- 9 Rice, A. M. *et al.* Evidence for Strong Mutation Bias toward, and Selection against, U Content in SARS-CoV-2: Implications for Vaccine Design. *Mol Biol Evol* **38**, 67-83, doi:10.1093/molbev/msaa188 (2021).
- 10 Mourier, T. *et al.* Host-directed editing of the SARS-CoV-2 genome. *Biochem Biophys Res Commun* **538**, 35-39, doi:10.1016/j.bbrc.2020.10.092 (2021).
- 11 Ruis, C. *et al.* Mutational spectra analysis reveals bacterial niche and transmission routes. *bioRxiv*, 2022.2007.2013.499881, doi:10.1101/2022.07.13.499881 (2022).
- 12 Laforge, M. *et al.* Tissue damage from neutrophil-induced oxidative stress in COVID-19. *Nat Rev Immunol* **20**, 515-516, doi:10.1038/s41577-020-0407-1 (2020).
- 13 Boers, J. E., Ambergen, A. W. & Thunnissen, F. B. Number and proliferation of basal and parabasal cells in normal human airway epithelium. *Am J Respir Crit Care Med* **157**, 2000-2006, doi:10.1164/ajrccm.157.6.9707011 (1998).
- 14 Boers, J. E., Ambergen, A. W. & Thunnissen, F. B. Number and proliferation of clara cells in normal human airway epithelium. *Am J Respir Crit Care Med* **159**, 1585-1591, doi:10.1164/ajrccm.159.5.9806044 (1999).
- 15 Meng, B. *et al.* Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature* **603**, 706-714, doi:10.1038/s41586-022-04474-x (2022).
- 16 Ruis, C. *et al.* Mutational spectra distinguish SARS-CoV-2 replication niches. *bioRxiv*, 2022.2009.2027.509649, doi:10.1101/2022.09.27.509649 (2022).
- 17 Paludan, S. R. & Mogensen, T. H. Innate immunological pathways in COVID-19 pathogenesis. *Sci Immunol* **7**, eabm5505, doi:10.1126/sciimmunol.abm5505 (2022).
- 18 Collier, A. Y. *et al.* Differential Kinetics of Immune Responses Elicited by Covid-19 Vaccines. *N Engl J Med* **385**, 2010-2012, doi:10.1056/NEJMc2115596 (2021).
- 19 Szczepanek, J. *et al.* Anti-SARS-CoV-2 IgG against the S Protein: A Comparison of BNT162b2, mRNA-1273, ChAdOx1 nCoV-2019 and Ad26.COV2.S Vaccines. *Vaccines (Basel)* **10**, doi:10.3390/vaccines10010099 (2022).
- 20 Klein, S. *et al.* SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun* **11**, 5885, doi:10.1038/s41467-020-19619-7 (2020).
- 21 Mendonca, L. *et al.* Correlative multi-scale cryo-imaging unveils SARS-CoV-2 assembly and egress. *Nat Commun* **12**, 4629, doi:10.1038/s41467-021-24887-y (2021).
- 22 Miller, S. & Krijnse-Locker, J. Modification of intracellular membrane structures for virus replication. *Nat Rev Microbiol* **6**, 363-374, doi:10.1038/nrmicro1890 (2008).
- 23 Turakhia, Y. *et al.* Ultrafast Sample placement on Existing tRees (USHER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nat Genet* **53**, 809-816, doi:10.1038/s41588-021-00862-7 (2021).
- 24 Khare, S. *et al.* GISAID's Role in Pandemic Response. *China CDC Wkly* **3**, 1049-1051, doi:10.46234/ccdcw2021.255 (2021).

Figures

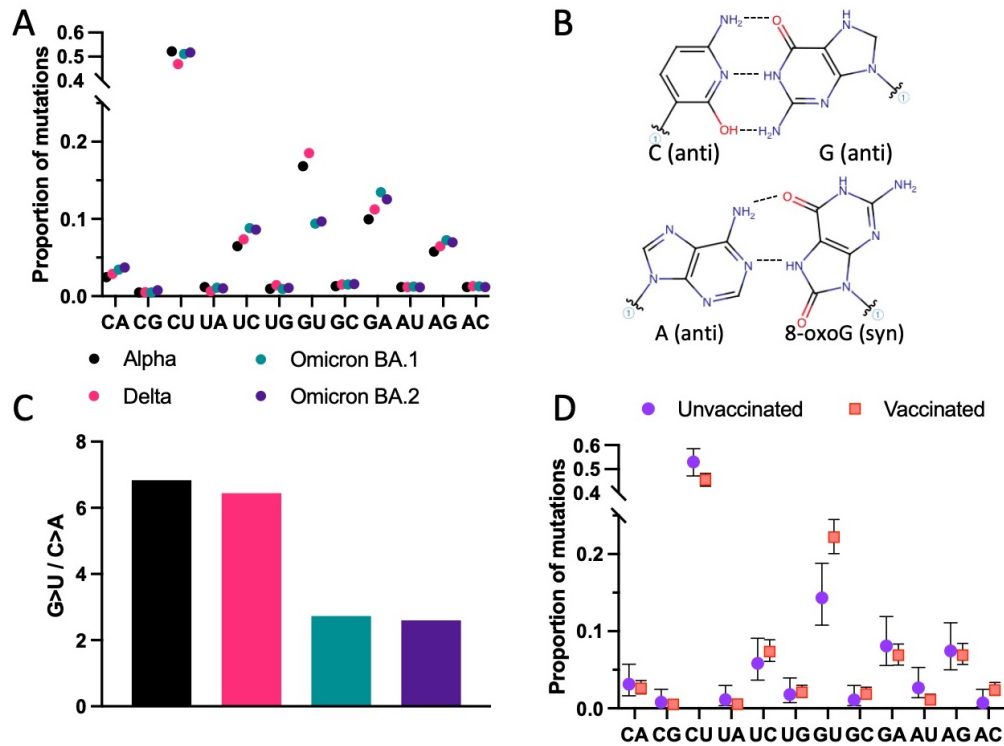


Fig.1. The environment of infected cells alters the G>U mutation incidence in SARS-CoV-2. (A) Analysis on mutation proportions across variants from ~180,000 samples collected worldwide before 15th of January 2021, when less than 0.5% of the world population was vaccinated. The different colours denote the variants. (B) Diagrammatic representation of standard (Watson and Crick) pairing of guanine and cytosine (top panel) and Hoogsteen base pairing between 8-oxoguanine and adenine (bottom panel). (C) Proportion of G>U and C>A within the variants, coloured as in (A). (D) Comparison of single base substitution spectra of unvaccinated and vaccinated patients. Error bars denote confidence intervals (CI); however, in (A), CI are smaller than the size of the points.

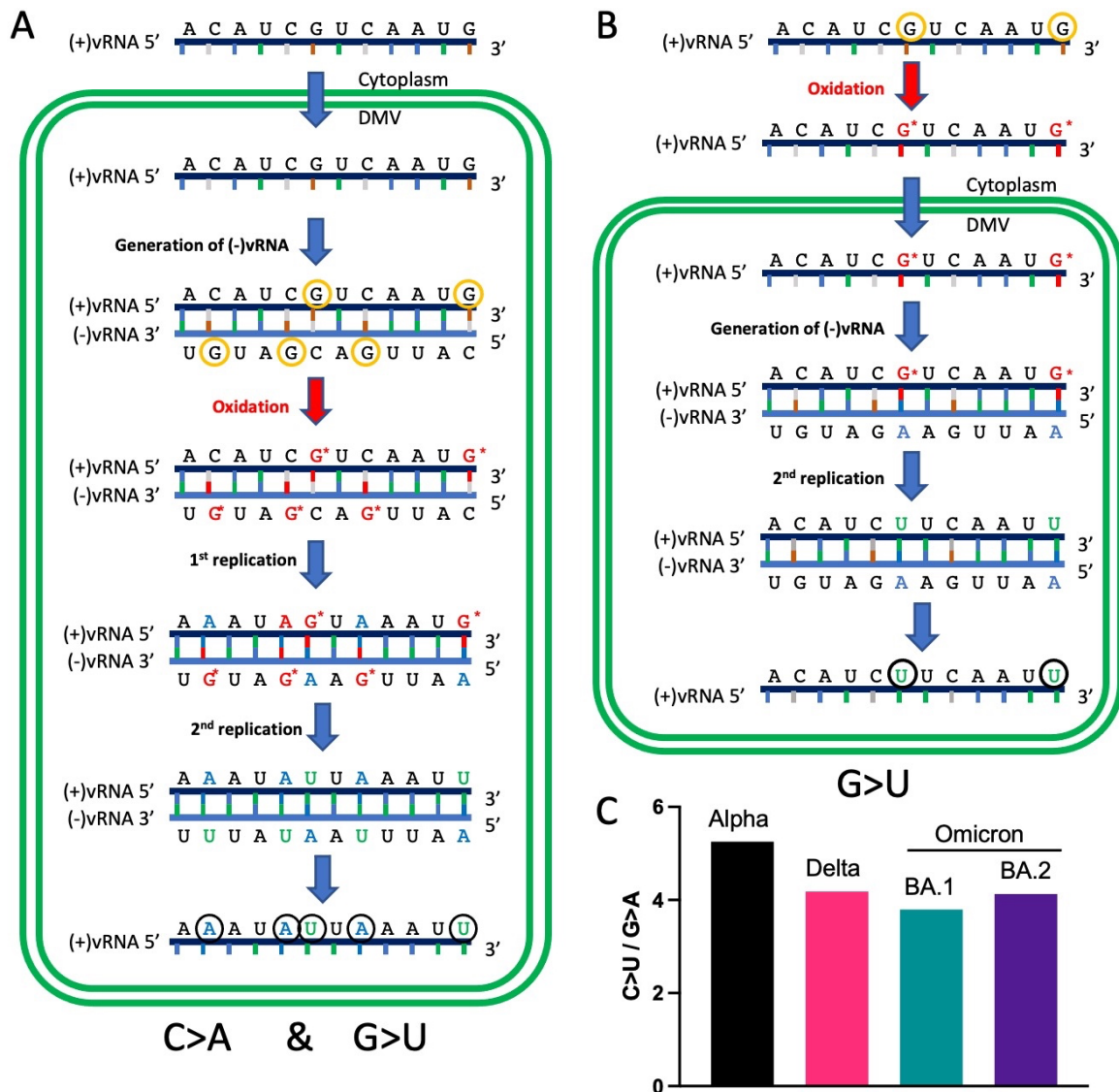


Fig.2. The influence of compartmentalisation on mutation patterns generated by oxidation. Differential pattern caused by mutagen exposure of vRNA guanines inside (A) or outside (B) double-membrane vesicles (DMVs), where vRNA replicates. Nucleotides circled in orange denote mutations that will occur in that scenario, while those in black mark the final product of the process. DMVs are delimited by double green lines. G* indicates 8-oxoguanine. (C) Proportion of C>U and G>A within the variants, coloured as in Fig.1A.