

Deciphering the evolutionary mechanisms of SARS-CoV-2; Absence of ORF8 protein and its potential advantage in the emergence of viral lineages

Lauro Velazquez-Salinas

Affiliation

Plum Island Animal Disease Center, Agricultural Research Service, USDA, Greenport, NY 11944, USA

Corresponding author:

Lauro Velazquez-Salinas: lauro.velazquez@usda.gov

To The Editor,

To date (2/5/2023), nearly three years after the official declaration of the COVID-19 pandemic, a total of 754, 018, 841 cases and 6,817,478 deaths have been officially reported to the world health organization (WHO) (<https://covid19.who.int/>). During this time, the global populace has witnessed the devastating effects of this pandemic, which can undoubtedly be considered the sanitary event of the century. In my opinion, understanding the mechanisms associated with the evolution of SARS-CoV-2 is the most important aspect of understanding the course of this pandemic. In this respect, the millions of viral sequences generated during the pandemic (1) have significantly aided in the tracking of the pandemic, resulting in the identification of multiple variants of concern.

At the start of the pandemic, the evolutionary patterns of SARS-CoV-2 and the potential role of natural selection in the emergence of new lineages was heavily debated (2). However, now several years into the pandemic, it is evident that SARS-CoV-2 has a tremendous ability to continuously evolve into new lineages with increased viral fitness (3) highlighting the significance of natural selection in this process (4, 5, 6).

Over the course of the pandemic, more than 1300 lineages have been identified under the Pango lineage classification (7). This situation has produced the diversification of this virus in multiple phylogenetic clades, which have been alternating their dominance in terms of their ability to rule the infections during this sanitary event (Figure 1A). Furthermore, multiple variants of concern, (VOC), and interest, (VOI), have been identified from these clades, with the clade GRA comprising the VOC omicron, which became dominant in 2022 (Figure 1B). In this sense, the role of natural selection in the evolutionary history of SARS-CoV-2 can be evidenced using multiple evolutionary algorithms based on the detection of natural selection at specific sites (MEME) or branch level (aBSREL and BUSTED) in a codon-based phylogenetic framework (8). As a result, the presence of multiple branches evolving under positive selection (Figure 1C) is consistent with the emergence and the dominance of different VOC (Figure 1B). Furthermore, the detection of these branches is associated with the discovery of specific residues under positive selection, which in some cases has been linked with increased infectivity (spike-452) (9), antibody resistance (spike-371) (10), generation of potential cleavage sites (spike-856) (11), transmissibility (Nucleocapsid-13) (12), and immune evasion (NSP6-105) (13), supporting the inferences

produced by multiple evolutionary algorithms, highlighting the relevance of natural selection during the evolution of this pathogen.

Conversely, when the strength of the natural selection was evaluated through the time of the evolution of different phylogenetic clades by contrasting evolutionary dN/dS (ratio of non-synonymous to synonymous substitutions) signatures between internal nodes and leaf nodes using the evolutionary algorithm Relax (14). Overall, strong evidence of positive selection was found at internal nodes in comparison to the leaf nodes (dN/dS= 1.5576 vs 0.5025 respectively). Evidence of relaxation of the natural selection at leaf nodes $K=0.73$ $p<0.05$ (K =relaxation/ intensification parameter) was inferred, with the diversifying selection, the main evolutionary force affected by the relaxed selection (Figure 1C). In light of these results, it is possible to suggest that the emergence of the lineages associated with different phylogenetic clades is initiated by a strong signal of diversifying selection (dN/dS= 411.436), which tends to relax during the time of circulation of different lineages. However, despite this relaxation in the diversifying selection, the strength of this evolutionary force may be considered strong enough (dN/dS= 82.176) to improve the viral fitness, leading to the emergence of new subvariants as is the case of the current dominant VOV omicron (15). Moreover, considering the current dominance of the VOC omicron and the lack of clinical cases produced by other VOC, the strength of the selection between GRA clade and the rest of the clades was compared (figure 1D). Interestingly, evidence of selection intensification was observed ($K=1.89$, $p<0.05$), affecting mainly the strength of diversifying selection during the evolution of lineages associated with GRA clade. The high difference in the strength of diversifying selection between GRA and the rest of the clades, clearly explains the dominance of the omicron variant over the rest of VOC. While more research is needed to understand the factors favoring the diversification of this variant, some evidence indicates that vaccine-breakthrough or antibody-resistant mutations may be the main mechanism associated with this process (16).

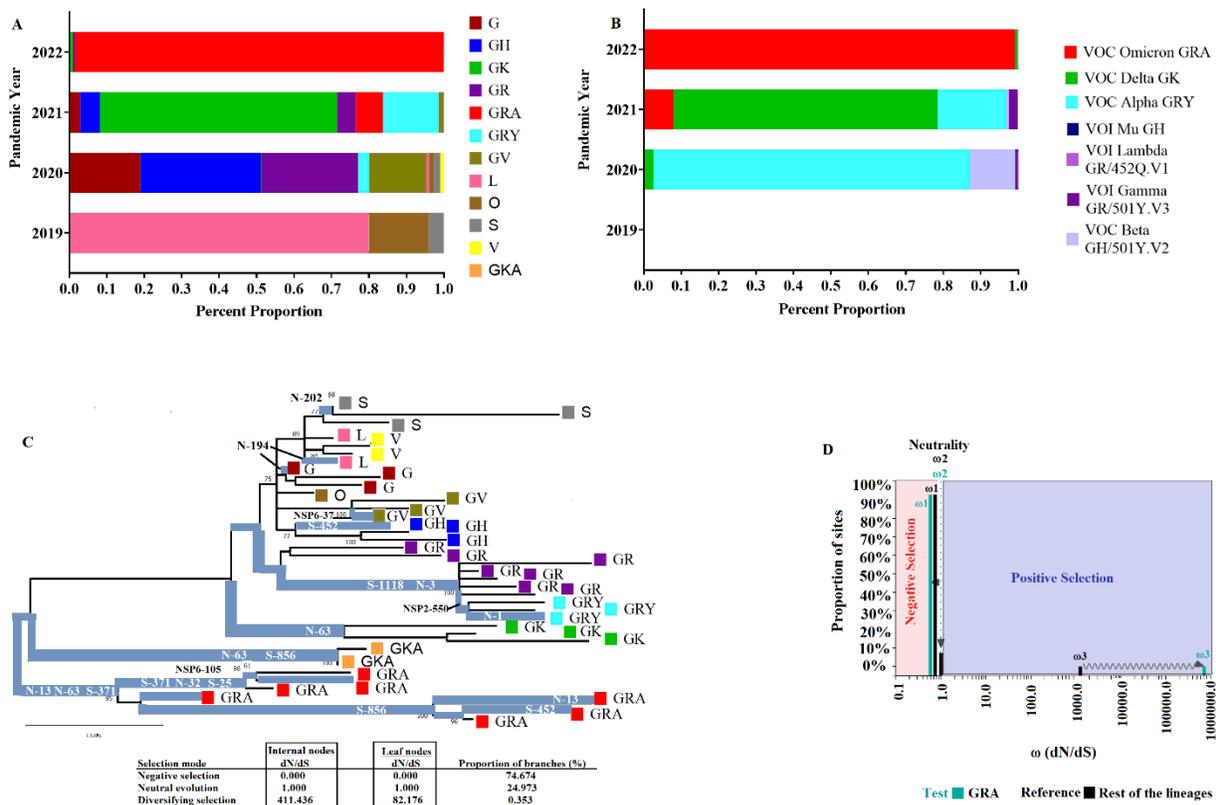


Figure 1. Phylogenomics of SARS-CoV-2 showing the relevance of diversifying selection during the evolution of this pathogen. A) Spatial dynamic of different phylogenetic clades during the evolution of SARS-CoV-2 (Based on GISAID classification). B) Circulation dynamic of diverse variants of concern (VOC) or variants of interest (VOI) during the pandemic. C) Phylogenetic analysis conducted by maximum likelihood using the general reversible model, showing multiple clades predicted during the evolution of SARS-CoV-2 (GISAID classification). Tree was constructed using 40 representative sequences. Determination of branches under diversifying selection was conducted using the algorithm aBSREL ($p < 0.0001$) (17) (blue branches with black asterisk). Evidence of diversifying selection on the tree was also confirmed by the algorithm BUSTED ($p < 0.0001$) (18). Sites under positive selection were determined by the algorithm MEME ($p < 0.05$) (19). The mapping of these sites at different branches in the tree was conducted by BUSTED considering an empirical bayes factor > 10 . Evidence of relaxation in the diversifying selection during the evolution of distinct lineages was produced by the comparison in the strength of the selection between internal nodes and leaf nodes using the algorithm RELAX. D) Evidence of intensified diversifying selection in the GRA clade (comprising the VOC omicron) in relation with the rest of the clades was obtained by the algorithm RELAX. Representative sequences and statistics about the number sequences associated with multiple clades, and VOC and VOI were obtained from the GISAID database (20).

However, beyond these evolutionary mechanisms, additional adaptive strategies should be considered as part of this complex evolutionary equation. In this regard, the article published by Colson et al., 2022 (21) is an interesting example of another potential evolutionary strategy used by SARS-CoV-2 to persist during this pandemic. In the study, the authors described the emergence of the subvariant Marseille-4B (B.1.160 Pangolin lineage; GH clade) circulating in France between September 2020 and March 2021. Remarkably, this variant carrying a nonsense mutation that inactivates the ORF8 gene, a protein that elicits a strong immune response during the infection of SARS-CoV-2 (22), became the dominant variant in December 2020 after the disappearance of multiple Marseille-A lineages. This observation led the authors to propose an intriguing hypothesis about the potential adaptive advantage of Marseille-4B after losing the ORF8 protein.

This hypothesis is supported based on the dN/dS values which suggest that the ORF8 gene may be subject to positive selection (Figure 2A). Furthermore, the presence of the predicted CTL and B cell epitopes (6, 23) in this protein raises the possibility that the absence of ORF8 might have helped Marseille-4B avoid both antibody and T cell recognition.

On the other hand, some important considerations regarding the function of ORF8 must be considered before supporting the hypothesis proposed by Colson et al. The ORF8 gene may also contribute to the antiviral properties of SARS-CoV-2, which is highlighted by its ability to downregulate the expression of MHC-I or lead to the inhibition of IFN β or IFN γ (24). This may explain why SARS-CoV-2 viruses lacking a functional ORF8 gene did not become the dominant variant during the pandemic, since the majority of these infections were associated with healthy immune competent individuals, a theory that is shared by others (27). Although the prevalence of viral lineages lacking ORF8 during the pandemic can be considered low, an increase of this phenotype was observed during 2020-2021 (Figure 2B), a situation mainly associated with the emergence of the VOC Alpha (Figures 2C and 2D). According to my estimation, the prevalence of omicron phenotypes lacking ORF8 was as low as 0.017%. However, considering the emergence of the Alpha variant and subvariant Marseille-4B, the lack of ORF8 protein in new emergent lineages should be monitored. Together this illustrates the importance of the hypothesis proposed by Colson et al. as a potential mechanism of adaptation of SARS-CoV-2. Future research is needed to understand the relevance of this finding.

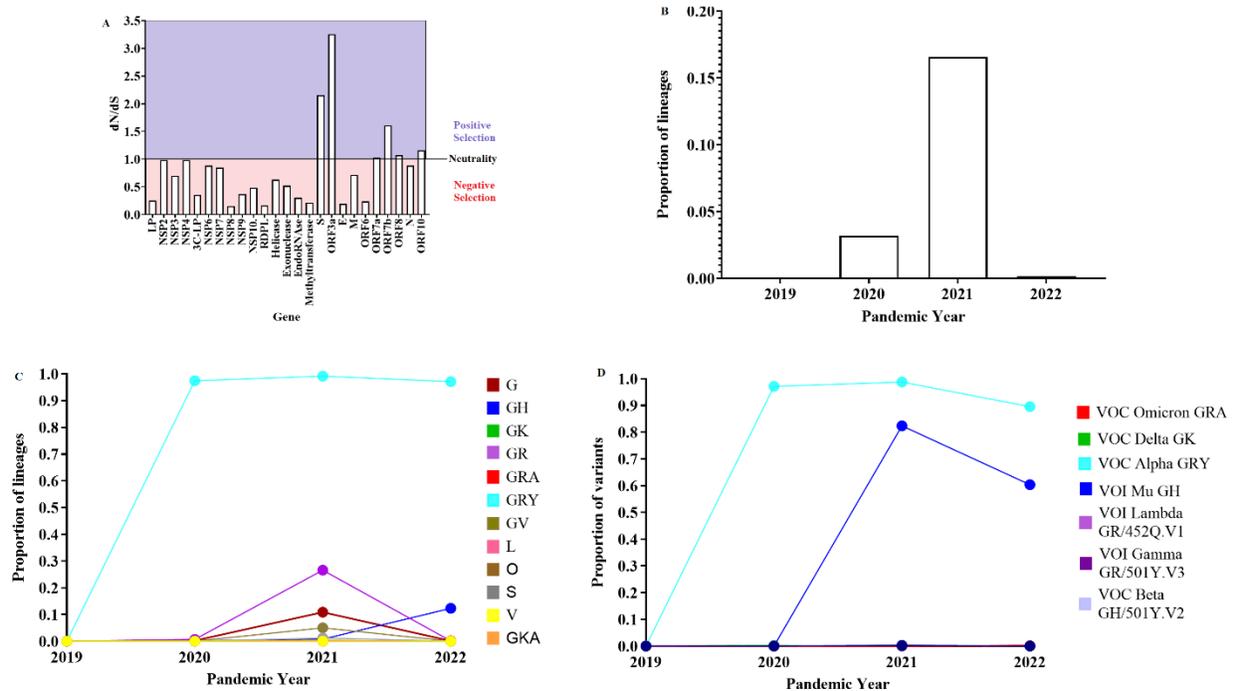


Figure 2. Prevalence of SARS-CoV-2 phenotypes lacking ORF8 protein. A) to put in context the evolutionary dynamic of ORF8 in comparison with the rest of the proteins present in SARS-CoV-2 during the pandemic, the sequences used to develop the tree shown in figure 1C were analyzed by the evolutionary algorithm SLAC (28). B) Using the database of GISAID, the overall proportion of lineages lacking ORF8 protein was quantified. Based on the information shown in figures 1A and 1B, the proportion of lineages (C) and VOC or VOI (D) lacking ORF8 protein in each phylogenetic clade was estimated.

Acknowledgments

The author thanks Dr. Lauren Holinka for her critical review of this letter.

Conflict of interest

Author declares that there is no conflict of interest

Funding

This work was conceived in absence of fundings.

Data Availability Statement

No datasets were generated during this work.

References

- 1- Maxmen, A., 2021. One million coronavirus sequences: popular genome site hits mega milestone. *Nature* 593, 21.
- 2- Mavian, C., Marini, S., Prosperi, M., Salemi, M., 2020. Authors' Reply to: Errors in Tracing Coronavirus SARS-CoV-2 Transmission Using a Maximum Likelihood Tree. Comment on "A Snapshot of SARS-CoV-2 Genome Availability up to April 2020 and its Implications: Data Analysis". *JMIR Public Health Surveill* 6, e24661.
- 3- Tegally, H., Moir, M., Everatt, J., Giovanetti, M., Scheepers, C., Wilkinson, E., Subramoney, K., Makatini, Z., Moyo, S., Amoako, D.G., Baxter, C., Althaus, C.L., Anyaneji, U.J., Kekana, D., Viana, R., Giandhari, J., Lessells, R.J., Maponga, T., Maruapula, D., Choga, W., Matshaba, M., Mbulawa, M.B., Msomi, N., consortium, N.-S., Naidoo, Y., Pillay, S., Sanko, T.J., San, J.E., Scott, L., Singh, L., Magini, N.A., Smith-Lawrence, P., Stevens, W., Dor, G., Tshiabuila, D., Wolter, N., Preiser, W., Treurnicht, F.K., Venter, M., Chiloane, G., McIntyre, C., O'Toole, A., Ruis, C., Peacock, T.P., Roemer, C., Kosakovsky Pond, S.L., Williamson, C., Pybus, O.G., Bhiman, J.N., Glass, A., Martin, D.P., Jackson, B., Rambaut, A., Laguda-Akingba, O., Gaseitsiwe, S., von Gottberg, A., de Oliveira, T., 2022. Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nat Med* 28, 1785-1790.
- 4- Velazquez-Salinas, L., Zarate, S., Eberl, S., Gladue, D.P., Novella, I., Borca, M.V., 2020. Positive Selection of ORF1ab, ORF3a, and ORF8 Genes Drives the Early Evolutionary Trends of SARS-CoV-2 During the 2020 COVID-19 Pandemic. *Frontiers in Microbiology* 11.
- 5- Martin, D.P., Weaver, S., Tegally, H., San, J.E., Shank, S.D., Wilkinson, E., Lucaci, A.G., Giandhari, J., Naidoo, S., Pillay, Y., Singh, L., Lessells, R.J., Ngs, S.A., UK, C.-G., Gupta, R.K., Wertheim, J.O., Nekturenko, A., Murrell, B., Harkins, G.W., Lemey, P., MacLean, O.A., Robertson, D.L., de Oliveira, T., Kosakovsky Pond, S.L., 2021. The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. *Cell* 184, 5189-5200 e5187.
- 6- Velazquez-Salinas, L., 2022. The complex evolutionary dynamics of SARS-CoV-2, a big challenge to control the pandemic of COVID-19. *J Med Virol* 94, 5082-5085.
- 7- O'Toole, A., Pybus, O.G., Abram, M.E., Kelly, E.J., Rambaut, A., 2022. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. *BMC Genomics* 23, 121.
- 8- Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V., Kosakovsky Pond, S.L., 2018. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol Biol Evol* 35, 773-777.
- 9- Zhang, Y., Zhang, T., Fang, Y., Liu, J., Ye, Q., Ding, L., 2022. SARS-CoV-2 spike L452R mutation increases Omicron variant fusogenicity and infectivity as well as host glycolysis. *Signal Transduct Target Ther* 7, 76.
- 10- Liu, L., Iketani, S., Guo, Y., Chan, J.F., Wang, M., Liu, L., Luo, Y., Chu, H., Huang, Y., Nair, M.S., Yu, J., Chik, K.K., Yuen, T.T., Yoon, C., To, K.K., Chen, H., Yin, M.T., Sobieszczyk, M.E., Huang, Y., Wang, H.H., Sheng, Z., Yuen, K.Y., Ho, D.D., 2022. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* 602, 676-681.

- 11- Maaroufi, H., 2022. The N764K and N856K mutations in SARS-CoV-2 Omicron S protein generate potential cleavage sites for SKI-1/S1P protease. *bioRxiv*, 2022.2001.2021.477298.
- 12- Oulas, A., Zanti, M., Tomazou, M., Zachariou, M., Minadakis, G., Bourdakou, M.M., Pavlidis, P., Spyrou, G.M., 2021. Generalized linear models provide a measure of virulence for specific mutations in SARS-CoV-2 strains. *PLoS One* 16, e0238665.
- 13- Benvenuto, D., Angeletti, S., Giovanetti, M., Bianchi, M., Pascarella, S., Cauda, R., Ciccozzi, M., Cassone, A., 2020. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. *J Infect* 81, e24-e27.
- 14- Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., Scheffler, K., 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol* 32, 820-832.
- 15- Hachmann, N.P., Miller, J., Collier, A.Y., Barouch, D.H., 2022. Neutralization Escape by SARS-CoV-2 Omicron Subvariant BA.4.6. *N Engl J Med* 387, 1904-1906.
- 16- Wang, R., Chen, J., Wei, G.W., 2021. Mechanisms of SARS-CoV-2 Evolution Revealing Vaccine-Resistant Mutations in Europe and America. *J Phys Chem Lett* 12, 11850-11857.
- 17- Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., Kosakovsky Pond, S.L., 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* 32, 1342-1353.
- 18- Murrell, B., Weaver, S., Smith, M.D., Wertheim, J.O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D.P., Smith, D.M., Scheffler, K., Kosakovsky Pond, S.L., 2015. Gene-wide identification of episodic selection. *Mol Biol Evol* 32, 1365-1371.
- 19- Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., Kosakovsky Pond, S.L., 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8, e1002764.
- 20- Khare, S., Gurry, C., Freitas, L., Schultz, M.B., Bach, G., Diallo, A., Akite, N., Ho, J., Lee, R.T., Yeo, W., Curation Team, G.C., Maurer-Stroh, S., 2021. GISAID's Role in Pandemic Response. *China CDC Wkly* 3, 1049-1051.
- 21- Colson, P., Gautret, P., Delerce, J., Chaudet, H., Pontarotti, P., Forterre, P., Tola, R., Bedotto, M., Delorme, L., Bader, W., Levasseur, A., Lagier, J.C., Million, M., Yahi, N., Fantini, J., La Scola, B., Fournier, P.E., Raoult, D., 2022. The emergence, spread and vanishing of a French SARS-CoV-2 variant exemplifies the fate of RNA virus epidemics and obeys the Mistigri rule. *J Med Virol*.
- 22- Hachim, A., Kavian, N., Cohen, C.A., Chin, A.W.H., Chu, D.K.W., Mok, C.K.P., Tsang, O.T.Y., Yeung, Y.C., Perera, R., Poon, L.L.M., Peiris, J.S.M., Valkenburg, S.A., 2020. ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. *Nat Immunol* 21, 1293-1301.
- 23- Moody, R., Wilson, K.L., Boer, J.C., Holien, J.K., Flanagan, K.L., Jaworowski, A., Plebanski, M., 2021. Predicted B Cell Epitopes Highlight the Potential for COVID-19 to Drive Self-Reactive Immunity. *Front Bioinform* 1, 709533.
- 24- Silvas, J.A., Vasquez, D.M., Park, J.G., Chiem, K., Allue-Guardia, A., Garcia-Vilanova, A., Platt, R.N., Miorin, L., Kehrer, T., Cupic, A., Gonzalez-Reiche, A.S., Bakel, H.V., Garcia-Sastre, A., Anderson, T., Torrelles, J.B., Ye, C., Martinez-Sobrido, L., 2021. Contribution of SARS-CoV-2 Accessory Proteins to Viral Pathogenicity in K18 Human ACE2 Transgenic Mice. *J Virol* 95, e0040221.

- 25- Li, J.Y., Liao, C.H., Wang, Q., Tan, Y.J., Luo, R., Qiu, Y., Ge, X.Y., 2020. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. *Virus Res* 286, 198074.
- 26- Geng, H., Subramanian, S., Wu, L., Bu, H.F., Wang, X., Du, C., De Plaen, I.G., Tan, X.D., 2021. SARS-CoV-2 ORF8 Forms Intracellular Aggregates and Inhibits IFN γ -Induced Antiviral Gene Expression in Human Lung Epithelial Cells. *Front Immunol* 12, 679482.
- 27- Chou, J.M., Tsai, J.L., Hung, J.N., Chen, I.H., Chen, S.T., Tsai, M.H., 2022. The ORF8 Protein of SARS-CoV-2 Modulates the Spike Protein and Its Implications in Viral Transmission. *Front Microbiol* 13, 883597.
- 28- Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22, 1208-1222.