



Parise Adadi ORCID iD: 0000-0003-4724-9463

Tarek Abd El-Aziz ORCID iD: 0000-0002-3441-9673

Murtaza TAMB UWALA ORCID iD: 0000-0001-8499-9891

Amos Lal ORCID iD: 0000-0002-0021-2033

Adam Brufsky ORCID iD: 0000-0001-8080-7960

Adam Brufsky ORCID ID 0000-0001-8080-7960

Letter to the Editor

Questions concerning the proximal origin of SARS-CoV-2

Murat Seyran^{1,2}, Damiano Pizzol^{3*}, Parise Adadi⁴, Tarek Mohamed Abd El-Aziz^{5,6},
Sk. Sarif Hassan⁷, Antonio Soares⁵, Ramesh Kandimalla^{8,9}, Kenneth Lundstrom¹⁰,
Murtaza Tambuwala¹¹, Alaa A. A. Aljabali¹², Amos Lal¹³, Gajendra Kumar Azad¹⁴,
Pabitra Pal Choudhury¹⁵, Vladimir N. Uversky¹⁶, Samendra P. Sherchan¹⁷,
Bruce D. Uhal¹⁸, Nima Rezaei^{19,20}, Adam M. Brufsky^{21*}

Affiliations: ¹Doctoral studies in natural and technical sciences (SPL 44), University of Vienna, Austria; ²Infection, Malignancy and Autoimmunity (NIIMA), Universal Scientific Education and Research Network (USERN), Austria; ³Italian Agency for Development Cooperation - Khartoum, Sudan Street 33, Al Amarat, Sudan; ⁴Department

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jmv.26478.

This article is protected by copyright. All rights reserved.

Accepted Article

of Food Science, University of Otago, Dunedin 9054, New Zealand; ⁵Department of Cellular and Integrative Physiology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Dr, San Antonio, TX 78229-3900, USA;

⁶Zoology Department, Faculty of Science, Minia University, El-Minia 61519, Egypt;

⁷Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram, Paschim Medinipur, 721140, West Bengal, India; ⁸CSIR-Indian Institute of Chemical Technology

Uppal Road, Tarnaka, Hyderabad-500007, Telangana State, India; ⁹Kakatiya Medical College/MGM-Hospital, DME/TSPSC, Hyderabad, Warangal, Telangana State-506007,

India; ¹⁰PanTherapeutics, Rte de Lavaux 49, CH1095 Lutry, Switzerland; ¹¹School of Pharmacy and Pharmaceutical Science, Ulster University, Coleraine BT52 1SA,

Northern Ireland, UK; ¹²Department of Pharmaceutical Sciences, Yarmouk University—Faculty of Pharmacy, Irbid 566, Jordan; ¹³Division of Pulmonary and Critical Care

Medicine, Mayo Clinic, Rochester, Minnesota, USA; ¹⁴Department of Zoology, Patna University, Patna-800005, Bihar, India; ¹⁵Applied Statistics Unit Indian Statistical

Institute, 203 B T Road, Kolkata 700108, West Bengal, India; ¹⁶Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612,

USA; ¹⁷Department of Environmental Health Sciences, Tulane University, New Orleans, LA, 70112, USA; ¹⁸Department of Physiology, Michigan State University, East Lansing, MI 48824, USA; ¹⁹Center for Immunodeficiencies, Pediatrics Center of Excellence,

Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran; ²⁰Network of Immunity in Infection, Malignancy and Autoimmunity (NIIMA), Universal

Scientific Education and Research Network (USERN), Tehran, Iran; ²¹Department of Medicine, Division of Hematology/Oncology, UPMC Hillman Cancer Center,

Pittsburgh, PA, USA

***Address correspondence to:** Damiano Pizzol, Italian Agency for Development Cooperation - Khartoum, Sudan Street 33, Al Amarat, Sudan. damianopizzol8@gmail.com; Adam Brufsky, MD, PhD, University of Pittsburgh School of Medicine, Division of Hematology-Oncology, 300 Halket Street, Pittsburgh, PA brufskyam@upmc.edu.

Keywords: Coronavirus, Genetic Variability, Mutation, Fusion Protein

Summary:

There is a consensus that Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) originated naturally from Bat coronaviruses (CoVs), in particular RaTG13. However, the SARS-CoV-2 host tropism/adaptation pattern has significant discrepancies compared to other CoVs, raising questions concerning the proximal origin of SARS-CoV-2. The flat and non-sunken surface of the sialic acid-binding domain of SARS-CoV-2 spike protein (S protein) conflicts with the general adaptation and survival pattern observed for all other CoVs. Unlike RaTG13, SARS-CoV-2 recombination presumably occurred between the S1/S2 domains of S protein enabling host furin protease utilization. Although millions of recorded cases have been recorded globally, SARS-CoV-2 S protein does not have any apparent further recombination, placing it in conflict with the recombination models of other CoVs. Similarly, the S protein receptor-binding domain (RBD) of SARS-CoV-2 has not accumulated high-frequency non-synonymous substitutions, differentiating SARS-CoV-2 from other CoVs that have positive selection/adaptation mutations in their RBDs.

Discussion:

Andersen and colleagues documented the possible natural origin of SARS-CoV-2 from BatCoV RaTG13 ^{1,2}. SARS-CoV-2 is the seventh zoonotic CoV virus capable of infecting humans, but the first and only human coronavirus (HCoV) with pandemic potential ³. Bat or rodent CoVs demonstrate certain specific changes in the S protein receptor-binding domain (RBD), as well as the S protein glycan-binding N-terminal

This article is protected by copyright. All rights reserved.

domain (NTD), during host tropism/adaptation ^{4,5}. SARS-CoV-2, unlike other CoVs, does not have those signature changes, suggesting that these RBD and NTD subdomains are of very recent origin.

The "Canyon Hypothesis" explains the development of canyons, depression zones or cavities on the surfaces of influenza virus, human rhinovirus, and Meningo viruses ⁶. In CoVs (except SARS-CoV-2), the S protein NTD domain has several predicted glycan-binding domains, with a common feature being the hidden localization of these glycan binding domains to cavities to limit their access to antibodies and immune cells ⁵. This pattern of CoVs is thought to be an evolutionary measure to restrict the recognition of these active sites by host immune system ⁴.

HCoVs can evade detection by host glycan-binding immune receptors. Comparative genomic analysis of six HCoVs with their corresponding native bat or rodent CoVs suggests compatibility with the "Canyon Hypothesis" resulting from various adaptive S protein NTD non-synonymous mutations near or at the glycan binding domain which are predicted to result in these NTD domains being hidden below the protein surface⁵. The predicted flat, non-sunken pattern of the SARS-CoV-2 S protein NTD glycan binding domains conflicts with this evolutionary host tropism/adaptation strategy ⁷.

A template-switching mechanism is presumably responsible for the high rate of RNA recombination in CoVs. In host cells, CoV RNAs show discontinuous RNA synthesis materialized by pauses of the RNA-dependent complex and subsequent jumps to downstream template acceptor sequences. This process results in subgenomic minus strand RNAs which serve as templates for subgenomic mRNAs. Due to the mechanistic similarity to recombination, this process might be at the origin of recombinant CoVs co-opting other CoV or even host related sequences⁸. Instances include the mouse hepatitis coronavirus S protein NTD sialic acid binding domain, likely arising from recombination of viral RNA with human galectin RNA sequences ⁸.

The furin recognition motif present at the SARS-CoV2 S1/S2 junction has no analogy in other "lineage B" beta-coronaviruses, including neither pangolin-CoV nor RaTG13¹. This indicates that the S protein S1/S2 junction is not a hot spot for RNA recombination termination that depends on a pattern swapping templates (copy-choice)⁸. Additionally, clinical isolates of SARS-CoV-2 S protein have not indicated any further recombination in this S1/S2 area, suggesting that the addition of a motif for S1/S2 site furin cleavage constituted a unique recombination occurrence. Finally, the CoV-unique insertion of 4 amino-acids creating a novel RRAR furin cleavage site introduces two arginine codons CGG-CGG, whose usage is extremely rare in CoVs, further supporting the hypothesis of a unique recombination occurrence.

HCoVs have high-frequency "hot spots" for non-synonymous amino acid replacements that can possibly create positive selection for host tropism/adaptation, resistance to neutralizing antibodies, or immune evasion². Interestingly, clinical SARS-CoV-2 isolates to date have only a single high frequency non-synonymous mutation, D614G, in their S protein⁹. Based on currently known mutation rates and patterns in clinical isolates of SARS-CoV-2, the S protein does not appear to be a mutational "hot spot" for SARS-CoV-2, unlike other human CoVs.

SARS-CoV-2 is the seventh HCoV, but the first HCoV with pandemic potential. SARS-CoV disappeared without a pandemic, and MERS-CoV is mostly endemic to the Arabian Peninsula with some additional limited traveler infections resulting in outbreaks in South Korea^{3,4}. These unique features of SARS-CoV-2 raise several questions concerning the proximal origin of the virus that require further discussion.

Author Contributions: M. S. conceived the study. A. M. B, V. N. U, A. L., and K. L. provided critical review. N. R., D. P., A. L. S. P. S., V. N. U, K. L.T. M. A. edited the article, M. S., P. A., T. M. A. formatted the article. All authors M. S., D. P., P. A., T. M.

This article is protected by copyright. All rights reserved.

A., S. S. H., A. S., R. K., K. L., D. M., M. T., A. L., P. P. C., V. N. U., S. P. S., B. D. U., N. R., A. M. B. and G. Z. K. of the consortium interpreted the results and approved the final version for submission.

Competing Interests statement: The authors declare no competing interests.

References

1. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med*. 2020;26(4):450-452. doi:10.1038/s41591-020-0820-9
2. Malaiyan J, Arumugam S, Mohan K, Gomathi Radhakrishnan G. An update on the origin of SARS-CoV-2: Despite closest identity, bat (RaTG13) and pangolin derived coronaviruses varied in the critical binding site and O-linked glycan residues [published online ahead of print, 2020 Jul 7]. *J Med Virol*. 2020;10.1002/jmv.26261. doi:10.1002/jmv.26261
3. Forni D, Cagliani R, Clerici M, Sironi M. Molecular Evolution of Human Coronavirus Genomes. *Trends Microbiol*. 2017;25(1):35-48. doi:10.1016/j.tim.2016.09.001
4. Hulswit RJ, de Haan CA, Bosch BJ. Coronavirus Spike Protein and Tropism Changes. *Adv Virus Res*. 2016;96:29-57. doi:10.1016/bs.aivir.2016.08.004
5. Li F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J Virol*. 2015;89(4):1954-1964. doi:10.1128/JVI.02615-14
6. Rossmann MG. The canyon hypothesis. Hiding the host cell receptor attachment site on a viral surface from immune surveillance. *J Biol Chem*. 1989;264(25):14587-14590.

This article is protected by copyright. All rights reserved.

7. Fantini J, Di Scala C, Chahinian H, Yahi N. Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. *Int J Antimicrob Agents*. 2020;55(5):105960. doi:10.1016/j.ijantimicag.2020.105960
8. Makino S, Keck JG, Stohlman SA, Lai MM. High-frequency RNA recombination of murine coronaviruses. *J Virol*. 1986;57(3):729-737. doi:10.1128/JVI.57.3.729-737.1986
9. Brufsky A. Distinct viral clades of SARS-CoV-2: Implications for modeling of viral spread [published online ahead of print, 2020 Apr 20]. *J Med Virol*. 2020;10.1002/jmv.25902. doi:10.1002/jmv.25902