SUPPLEMENTARY MATERIALS
to the article by A.Yu. Palyanov, A.P. Devyaterikov, N.V. Palyanova, A.M. Shestopalov
"Evolutionary inferences from the analysis of mutation dynamics
in the SARS-CoV-2 replication-transcription complex"

Supplementary Material 1. Comparison of SARS-CoV-2 reference genome (Wuhan/WIV04/2019) with other most similar viruses, with the focus on RTC proteins

Table S1. Comparison of SARS-CoV-2 Wuhan/WIV04/2019 proteins nsp7–nsp14 with most similar viruses RaTG13 and BANAL-52, and with the typical SARS-CoV-2 descendant (Omicron) after almost 4 years of evolution

Virus name	hCoV-19/ Wuhan/ WIV04/2019	hCoV-19/bat/ Laos/ BANAL-52/2020	BetaCoV/bat/ Yunnan/ RaTG13/2013	hCoV-19/ Netherlands/ UT-RIVM-134727/2023
GISAID EPI ISL	402124	4302644	402131	18674579
Collection date	30.12.2019	05.07.2020	24.07.2013	07.10.2023
Genome length, nt	29,903	29,838	29,855	29,721
Entire genome nucleotide identity with hCoV-19/ Wuhan/WIV04/2019	–	96.8 %	96.1 %	99.43 %
nsp7 length (aa) / substitutions	83	N69S	–	–
nsp8 length/ substitutions	198	–	–	–
nsp8 length/ substitutions	113	–	–	(T35I)
nsp8 length/ substitutions	139	–	–	–
nsp8 length/ substitutions	14	–	–	–
nsp8 length/ substitutions	932	N198D	K143R, N198D, V233I, Y719H	D63N, G671S, P323L
nsp8 length/ substitutions	601	–	A505T	S36P
nsp8 length/ substitutions	527	P203L, W348C	I80V, R212K, V101I	I42V
Total nsp7–nsp14 length / N and % of substitutions	2,607	4 (0.15 %)	8 (0.30 %)	7 (0.27 %)

Note. The table contains only substitutions because three genomes (columns 3, 4, and 5) turned out to have no deletions or insertions in the nsp7–nsp14 genes relative to the SARS-CoV-2 reference genome (Wuhan/WIV04/2019, column 2).

Supplementary Material 2. The data illustrating the emergence of the first cases of SARS-CoV-2 genome variants with the nsp12:G671S mutation

Table S2. Chronology of the emergence of SARS-CoV-2 genome variants with the nsp12:G671S mutation

Date	GISAID EPI ISL	Clade	Pango lineage	Location
11.03.2020	852796	B.1	G	Europe / Germany / Baden-Wurttemberg / Freiburg
14.03.2020	852785	B.1	G	Europe / Germany / Baden-Wurttemberg / Freiburg
15.03.2020	603845	B.1	GH	North America / USA / Utah
16.03.2020	456138	B.1	G	South America / Colombia / Valle del Cauca / Cali
16.03.2020	852781	B.1	G	Europe / Germany / Baden-Wurttemberg / Freiburg
24.03.2020	571214	B.1	GH	North America / USA / Michigan
31.03.2020	447755	B.1	G	South America / Colombia / Valle del Cauca / Buga
05.04.2020	11254235	B.1	GH	North America / USA / Illinois / Cook County
05.04.2020	819318	B.1	G	Europe / Spain / Catalunya
07.04.2020	12416249	B.1	GH	North America / USA / Maryland
08.04.2020	11254307	O	A	North America / USA / Illinois / Chicago
09.04.2020	11254331	B.1	GH	North America / USA / Illinois / Chicago
09.04.2020	511903	B.1.206	G	Asia / India / West Bengal / Kolkata
10.04.2020	514474	B.1.1	GR	Europe / United Kingdom / England
10.04.2020	11254232	B.1	GH	North America / USA / Illinois / Cook County
11.04.2020	8482147	B.1	GH	North America / USA / Illinois / Chicago
12.04.2020	11254293	B.1	G	North America / USA / Illinois / Chicago
13.04.2020	11254292	S	A	North America / USA / Illinois / Chicago
14.04.2020	508339	B.1	G	Asia / India / West Bengal / Kolkata
20.04.2020	455648	B.1	G	Asia / India / West Bengal / Kolkata
20.04.2020	11254316	B.1	G	North America / USA / Illinois / Chicago
20.04.2020	508341	B.1	G	Asia / India / West Bengal / Kolkata
21.04.2020	876811	B.1.1	GR	Europe / Italy / Bari / Adelfia
22.04.2020	11254350	B.1	G	North America / USA / Illinois / Cook County
22.04.2020	486866	B.1.1	GR	Africa / Senegal / Dakar
23.04.2020	11254306	B.1	G	North America / USA / Illinois / Chicago
23.04.2020	486864	B.1.1	GR	Africa / Senegal / Dakar
24.04.2020	11254332	B.1	G	North America / USA / Illinois / Chicago
25.04.2020	2885131	S	A.3	North America / USA / Maryland
25.04.2020	508353	B.1	G	Asia / India / West Bengal / Kolkata
26.04.2020	19537034	B.1.617.2	GV	Europe / United Kingdom

Note. The contents of the table, including the collection date, genome's GISAID ID, clade, Pango lineage, and geographic location, are entirely based on the genomic and metagenomic data available in the GISAID database. Only high-quality genomes were considered (filtered using the *complete*, *low coverage excluded*, and *collection date complete* options).

According to Table 2 and data from GISAID, the first genome variant with the nsp12:G671S mutation was detected on March 11, 2020, approximately one year before the increase in its population frequency began. This sample originated from Germany and belonged to clade B.1 and Pango lineage G. During the following month, the mutation was observed mainly in Germany and the United States, where it was first detected on March 15, 2020. On April 21, 2020, the first genome of B.1.1/GR with nsp12:G671S was detected in Italy, and on April 26, 2020, the first case of B.1.617.2, better known as Delta, was registered in the UK. According to GISAID data, the first case of B.1/G was on January 1, 2020, in Niger; B.1.1/GR on February 3, 2020, in Western Russia; and the first case of B.1.617.2 April 26, 2020, in the UK. Interestingly, this first case of Delta contained the nsp12:G671S mutation.

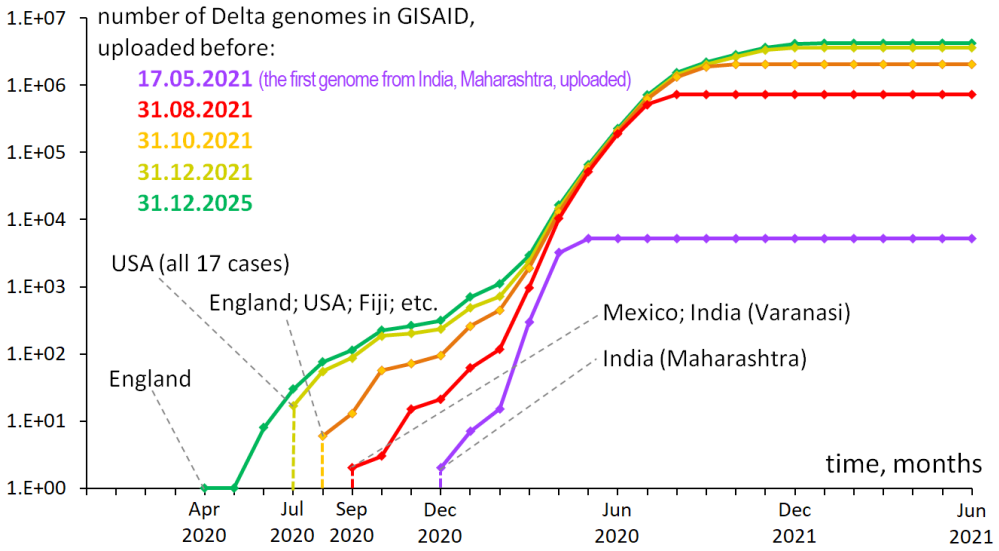

Fig. S1. Effect of submission date cutoff on apparent emergence of Delta variant in GISAID.

Cumulative number of genomes labeled as "B.1.617.2+AY.*(Delta)" in GISAID as a function of their collection date. Each curve corresponds to a different submission-date cutoff, with color indicating the latest included submission date (green represents the current dataset). As submission-date threshold increases, apparent earliest occurrences of Delta genomes shift backwards in time.

Figure S1 illustrates how delayed data deposition can substantially affect the perceived timeline of variant emergence. In GISAID, intervals of one to three years between collection and submission are common, meaning that the full retrospective picture of the pandemic becomes visible only long after the events themselves.

Supplementary Material 3. Details of mutational dynamics during the 2nd tide of nsp12:G671S

The monthly dynamics of the number of collected samples from December 2021 to February 2023 is shown in Figure S2A, broken by geographical regions. The interval was chosen to cover the end of the first period of nsp12:G671S dominance and the beginning of the second one, allowing a detailed analysis of the events between these phases.

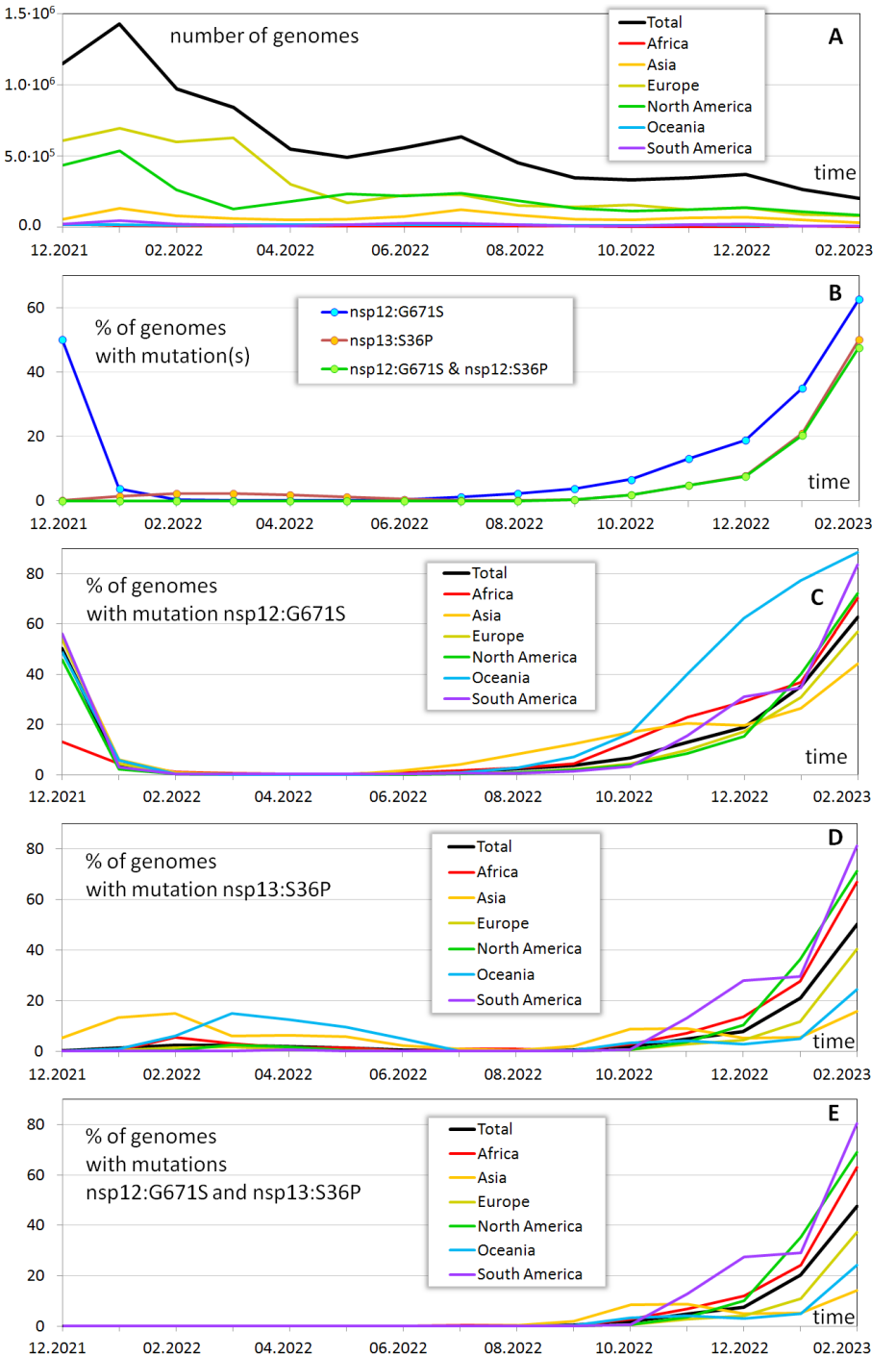

Fig. S2. The effect of the emergence of genomes simultaneously containing nsp12:G671S and nsp13:S36P on the dynamics of their prevalence in the population.

Monthly dynamics of (A) the total number of genomes and their numbers by location, (B) percent of genomes with mutation nsp12:G671S, nsp13:S36P, and both at once, (C) percent of genomes with mutation nsp12:G671S by location, (D) percent of genomes with mutation nsp13:S36P by location, and (E) percent of genomes with both mutations at once by location. Time range: from December 2021 to February 2023.

In December 2021, nsp13:S36P was almost absent, while the prevalence of nsp12:G671S was declining, accounting for approximately 50 % of all samples (Fig. S2B). At the beginning of 2022, nsp13:S36P began to grow: in February 2022, a local maximum of 18 % was observed in Asia, followed by a decline to zero by July 2022. A similar pattern occurred in Oceania, where a local maximum of 18 % in February 2022 was also followed by a complete decline by July 2022 (Fig. S2D). Overall, this contributed little to the global population on average: the share of samples with nsp13:S36P did not exceed 2.5 % until July 2022. The minimal, nearly zero prevalence of nsp12:G671S mutation was recorded in April–May 2022, and that of nsp13:S36P in July–August 2022 (Figs. S2C, D). Until this point, the two mutations were rarely found together within the same genomes (Fig. S2E).

The situation began to change significantly from September 2022 onward (Fig. S2E) with a rapid increase in the number of genomes carrying both nsp12:G671S and nsp13:S36P reaching an average of 35 % by February 2023, and 60–80 % in certain regions. However, this growth was preceded by an earlier process starting around July 2022 – an increase in the presence of samples with nsp12:G671S in Asia and Oceania locations (which were leading between February and April 2022 as well) – see Fig. S2C. In most cases, nsp13:S36P and nsp12:G671S appeared together, whereas nsp13:S36P alone was rarely detected. On the contrary, nsp12:G671S without nsp13:S36P dominated from July to November 2022. By December 2022, the percentages of nsp12:G671S with and without nsp13:S36P were approximately equal, and then the proportion of genomes carrying nsp12:G671S alone continued to decline.

Thus, both the decline to zero and the subsequent rise in the proportion of genomes carrying nsp12:G671S began about three months earlier than the corresponding increase in nsp13:S36P. The earliest rise in the percentage of genomes simultaneously including both nsp12:G671S and nsp13:S36P was observed in Asia. After 1–2 months, similar increases appeared in other regions, reaching levels of 60–80 % in some locations by January 2023. Based on these observations, two hypotheses can be proposed:

(1) Nsp12:G671S first emerged in Asia in July 2022 and, by August 2022, had already spread and been detected across all major geographical regions. By February 2023, the proportion of samples with this mutation exceeded 60 % on average worldwide, and in South America and Oceania, 80 %. The nsp13:S36P mutation appeared among samples with nsp12:G671S in September–October 2022, also in Asia. There, its spread remained limited – the proportion of such genomes did not exceed 20 % by February 2023. However, after reaching other regions, likely more favorable for its dissemination, its frequency increased rapidly, reaching an average of 50 % globally and over 90 % in South America by February 2023.

(2) Nsp12:G671S and, later, nsp13:S36P (in genomes already carrying nsp12:G671S), arose independently on all continents. This may reflect an altered viral fitness landscape shaped by changes in population immunity, the presence of competing variants, and possibly other contributing factors. By the time the nsp13:S36P mutation appeared, changes caused by the emergence of the nsp12:G671S mutation (probably advantageous to variants with nsp13:S36P) were added. As a result of this combination of factors, genomes containing nsp12:G671S alone or together with nsp13:S36P appear to have achieved higher fitness than other existing or potential variants, which explains their repeated emergence and global spread. The probability of a single nucleotide substitution occurring independently is not negligible; such events happen continuously, but only those that confer a selective advantage persist and become epidemiologically visible.

Supplementary Material 4. Literature search results on the effects of mutations listed in Table 3

Nsp9:T35I: "Structural analysis showed that nsp9:T35I substitution may affect the structure of the replication and transcription complex (RTC), which might in turn affect the RNA synthesis of the virus" (Ip et al., 2024).

Nsp12:P323L, Nsp12:G671S: "SARS-CoV-2 variants with NSP12 P323L/G671S mutations display enhanced virus replication in ferret upper airways and higher transmissibility" (Kim et al., 2023).

Nsp12:G671S: "~97.8 % of Delta and ~99 % of XBB Omicron subvariant harbor an additional G671S mutation in NSP12" and "Structural and MD simulation analysis of the SARS-CoV-2 NSP12-NSP8 complex showed that NSP12 P323L/G671S mutations reinforce the interface interaction between NSP12 and NSP8 through strengthened hydrophobic interactions, resulting in improved assembly of the replication machinery for efficient viral replication and transmission. This hypothesis was substantiated by MST and pull-down assays to verify enhanced NSP12-NSP8 complex interaction. Thus, compared to NSP12 WT, the NSP12 P323L and P323L-G671S mutants have significantly increased interaction with NSP8, resulting in enhanced RdRp activity" (Kim et al., 2023).

"Position 15,451 (nsp12:671) has been extensively studied, as the G15451A (nsp12:G671S) mutation is a defining mutation for many Delta and Omicron lineages. Pitts et al. explicitly tested the potency of remdesivir against SARS-CoV-2 viruses harboring either the nsp12:P323L mutation or the nsp12:P323L/G671S double mutation and found no significant difference in the drug's EC50 (35). Kim et al. explored this further, finding that the P323L/G671S double mutation enhances nsp12-nsp8 binding and increases overall replicative fitness of the virus in vivo (60)" (Ling-Hu et al., 2025).

"P323L and G671S are located near the NSP12 active site" and "In summary, two of the top nonsynonymous mutations under strong beneficial selection (P323L and G671S) were confirmed to impact NSP12 fitness by conferring replicative advantages, stabilized replication assembly and viral RNA processing, and modulation of drug potencies" (Paradis, Wu, 2024).

Nsp12:Y273H, Nsp12:P323L, Nsp12:G671S: "Remdesivir retained activity against all of the major SARS-CoV-2 variants, suggesting the RdRp substitutions Y273H, P323L, and G671S do not affect the compound" (Iketani, Ho, 2024).

Nsp12:D284Y: "The D284Y mutation may affect the stability and flexibility of the Nsp12 protein. This mutation is a characteristic of the SARS-CoV-2 variant NB.1.8.1" (Kumar et al., 2022).

Nsp12:L838I: "The more complex mutational profile P323L+G671S+L838I+D738Y+K91E, which was found with a prevalence of 2.6 %, showed a delayed reduced response to remdesivir, as confirmed by the increase in SARS-CoV-2 viral load and by a reduced theoretical binding affinity versus RdRp" (Gratteri et al., 2023). We revealed that the nsp12:D738Y and nsp12:K91E mutations are found in GISAID in quantities of 6,860 and 7,068 out of 15.5 million, which is significantly less than 1 %.

Nsp13:S36P: "When we searched for NSP13 mutations in major Omicron subvariants, we saw that NSP13-S36P emerged in XBB.1. However, no experimental studies have been performed on the functional significance of this specific mutation" (Furusawa et al., 2025).

Nsp13:P77L: "The Delta clades also share mutations in the ORF7a (V82A and T120I) and nsp13 (P77L) proteins involved in innate-immune antagonization" (Bloom et al., 2023).

Nsp13:M233I: "The results show that the nsp13:M233I and nsp14:D222Y substitutions reduce viral polymerase activity" (Furusawa et al., 2025).

Nsp13:N268S, Nsp12:Y273H: "For the nsp12:Y273H test, among the BQ1 samples, only one was positive. In contrast, for the nsp13:N268S test, three out of the nine BQ1 samples were found to be positive, indicating a probable probe mismatch or suboptimal conditions for the assay" (Zniber et al., 2024).

Nsp13:R392C: "The R392C mutation is located close to three of those six residues (Asp374, Glu375, Gln404). Thus, the mutation might impact the NTPase activity of NSP13. Considering NSP13 as the potent antiviral drug target, this mutation might be crucial in interfering with one of the fragments to nearly resided Arg382" (Hossain et al., 2022).

Nsp14:I42V: (1) "In short, our results identified nsp14:I42V, E:T9I, and M:(D3G/Q19E/A63T) mutations in combination significantly reduced the neurovirulence of SARS-CoV-2 in K18-hACE2 mice" (Sangare et al., 2025) and (2) "One NSP14 mutation, I42V, was fixed in BA.1 and all subsequent Omicron subvariants. … Yet, there has been only one fixed mutation in NSP14 after five years. Therefore, degeneration of non-structural genes may be more pronounced than what we observe in the lineage-defining mutation spectrums" (Liu, 2025).

Nsp14:A394V: “Specifically, the delta variant possessed the NSP14 A394V mutation, which occurs in ~30 % of total GISAID-archived variants. However, direct comparison of mutation frequency across distinct delta variant sequences did not support the conclusion that the A394V mutation conferred a reduction in enzyme function” (Shakran et al., 2022).

References

Bloom J.D., Beichman A.C., Neher R.A., Harris K. Evolution of the SARS-CoV-2 mutational spectrum. *Mol Biol Evol.* 2023;40(4): msad085. doi 10.1093/molbev/msad085

Furusawa Y., Kiso M., Uraki R., Sakai-Tagawa Y., Nagai H., Koga M., Kashima Y., … Yotsuyanagi H., Halfmann P.J., Kamitani W., Yamayoshi S., Kawaoka Y. Amino acid substitutions in NSP6 and NSP13 of SARS-CoV-2 contribute to superior virus growth at low temperatures. *J Virol.* 2025;99(3):e0221724. doi 10.1128/jvi.02217-24

Gratteri C., Ambrosio F.A., Lupia A., Moraca F., Catalanotti B., Costa G., Bellocchi M., … Sarmati L., Svicher V., Bryant S., Artese A., Alcaro S. Molecular and structural aspects of clinically relevant mutations of SARS-CoV-2 RNA-dependent RNA polymerase in remdesivir-treated patients. *Pharmaceuticals*. 2023;16(8):1143. doi 10.3390/ph16081143

Hossain A., Akter S., Rashid A.A., Khair S., Alam A.S.M.R.U. Unique mutations in SARS-CoV-2 Omicron subvariants’ non-spike proteins: potential impacts on viral pathogenesis and host immune evasion. *Microb Pathog.* 2022;170:105699. doi 10.1016/j.micpath.2022.105699

Iketani S., Ho D.D. SARS-CoV-2 resistance to monoclonal antibodies and small-molecule drugs. *Cell Chem Biol.* 2024;31(4):632-657. doi 10.1016/j.chembiol.2024.03.008

Ip J.D., Chu W.M., Chan W.M., Chu A.W., Leung R.C., Peng Q., Tam A.R., … Yuen K.Y., Kok K.H., Shi Y., Hung I.F., To K.K. The significance of recurrent *de novo* amino acid substitutions that emerged during chronic SARS-CoV-2 infection: an observational study. *EBioMedicine.* 2024;107:105273. doi 10.1016/j.ebiom.2024.105273

Kim S.M., Kim E.H., Casel M.A.B., Kim Y.I., Sun R., Kwak M.J., Yoo J.S., … Hwang J., Song M.S., Kim M.H., Jung J.U., Choi Y.K. SARS-CoV-2 variants with NSP12 P323L/G671S mutations display enhanced virus replication in ferret upper airways and higher transmissibility. *Cell Rep.* 2023;42(9):113077. doi 10.1016/j.celrep.2023.113077

Kumar S., Kumari K., Azad G.K. Emerging genetic diversity of SARS-CoV-2 RNA dependent RNA polymerase (RdRp) alters its B-cell epitopes. *Biologicals.* 2022;75:29-36. doi 10.1016/j.biologicals.2021.11.002

Ling-Hu T., Simons L.M., Rios-Guzman E., Carvalho A.M., Agnes M.F.R., Alisoltanidehkordi A., Ozer E.A., Lorenzo-Redondo R., Hultquist J.F. The impact of remdesivir on SARS-CoV-2 evolution *in vivo*. *JCI Insight*. 2025;10(4):e182376. doi 10.1172/jci.insight.182376

Liu Y. Is SARS-CoV-2 facing constraints in its adaptive evolution? *Biomol Biomed*. 2025;25(11):2407-2415. doi 10.17305/bb.2025.12537

Paradis N.J., Wu C. Enhanced detection and molecular modeling of adaptive mutations in SARS-CoV-2 coding and non-coding regions using the c/μ test. *Virus Evol.* 2024;10(1):veae089. doi 10.1093/ve/veae089

Sangare K., Liu S., Selvaraj P., Stauft C.B., Starost M.F., Wang T.T. Combined mutations in nonstructural protein 14, envelope, and membrane proteins mitigate the neuropathogenicity of SARS-CoV-2 Omicron BA.1 in K18-hACE2 mice. *mSphere.* 2025;10(1): e00726-24. doi 10.1128/msphere.00726-24

Shakran D.A., Mikbel D.M., Vilela M.F., Benoit L.A. Memory gaps in America: mutational and immunoinformatic analysis of evolving SARS-CoV-2 variants of concern and interest. *ImmunoHorizons.* 2022;6(1):1-7. doi 10.4049/immunohorizons.2100096

Zniber N.M., Elbenaissi Y., Aadi Y., Elkochri S., Tagajdid M.R., Agoudim S., Elannaz H., Laraqui A., Elmchichi B., Touil N., Ennibi K., Amine I.L., Abi R The rapid screening of SARS-CoV-2 variants can be an alternative solution for low-income countries? *Clin Case Rep Int.* 2024;8(1):1697. doi 10.25107/2638-4558.1697