**Supporting Information**

**Emerging SARS-CoV-2 Lineages in Middle Eastern Jordan with Increasing Mutations Near Antibody Recognition Sites**

Rima Hajjo, Dima A. Sabbah, Sanaa K. Bardaweel

**Affiliations:**

Department of Pharmacy, Faculty of Pharmacy, Al-Zaytoonah University of Jordan, Amman, Jordan (R.H., D.S.)

Department of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA (R.H.)

National Center for Epidemics and Communicable Disease Control, Amman, Jordan (R.H.)

Department of Pharmaceutical Sciences, School of Pharmacy, University of Jordan, Amman, Jordan (S.B.)

**Correspondence:**

Rima Hajjo, Department of Pharmacy, Faculty of Pharmacy, Al-Zaytoonah University of Jordan, P.O. Box 130 Amman 11733 Jordan.

[rhajjo@gmail.com](mailto:rhajjo@gmail.com); [r.hajjo@zuj.edu.jo](mailto:r.hajjo@zuj.edu.jo)

**Materials and Methods**

**Epidemiology and genomic data.** COVID-19 epidemiology data have been retrieved from the WHO COVID-19 dashboard(1) and the Jordanian Ministry of health website(2). All 556 genomic sequences for SARS-CoV-2 viruses found in Jordan were downloaded from GISAID(3).

**Genomic analysis.** CoV-GLUE was used for the analysis and interpretation of SARS-CoV-2 virus genomic sequences, with a focus on amino acid sequence variation. This freely available web application is based on the GLUE(5) environment and provides a browsable database of amino acid replacements and coding region indels that have been observed in sequences from the COVID-19 pandemic. Pair-wise alignments were also performed for all was performed to all SARS-CoV-2 virus sequences of interest using MAFFT(6). After excluding irrelevant sequences and those with potential quality issues, amino acid replacements and in-frame indels in each sequence were identified.

**Mutational analysis.** CoVsurver(6) from GISAID was used in order to identify to identify candidates for phenotypic changes of special epidemiological relevance. Additionally, this application enabled the visualization of the position of the relevant mutation(s) in structural models and highlighted whether these mutations are close to common host receptor or antibody binding sites.

**Phylogenetic assignment of named global outbreak lineages.** Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN)(7) version v3.0 was used to assign lineages to query sequences downloaded from GISAID. This software was developed by the Centre for Genomic Pathogen Surveillance. All methods implemented in this software are described in Rambaut *et al* 2020(8).

**Clade and lineage nomenclature.** We used GISAID’s nomenclature system for classifying major clades. Clade assignments are based on genetic markers (i.e., marker mutations) within seven high-level phylogenetic groupings from the early split of S and L clades, to the further evolution of L into V and G and later of G into GH, GR and GV (cf. Table 1). Additionally, GISAID clades were also described in details based on lineages assigned by PANGOLIN, which allows for a better understanding of the global spread of COVID-19.

**Data visualization.** Jalview(9) v.2.11.1.3, CoVsurverand COVID CoV Genomics(10) tools were used to analyze viral sequences and to identify mutations from the multiple sequence alignment results.

**References**

1.         WHO Coronavirus Disease (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard [Internet]. [cited 2020 Oct 27]. Available from: https://covid19.who.int/

2.         Ministry of Health, the official website of the Jordanian Ministry of Health | Coronavirus disease [Internet]. [cited 2021 Jan 28]. Available from: https://corona.moh.gov.jo/en

3.         Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality [Internet]. Vol. 22, Eurosurveillance. European Centre for Disease Prevention and Control (ECDC); 2017 [cited 2021 Feb 1]. p. 1. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5388101/

4.         Singer JB, Thomson EC, McLauchlan J, Hughes J, Gifford RJ. GLUE: A flexible software system for virus sequence data. BMC Bioinformatics [Internet]. 2018 Dec 18 [cited 2021 Feb 9];19(1):532. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2459-9

5.         Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution [Internet]. 2013 Apr 1 [cited 2021 Feb 9];30(4):772–80. Available from: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst010

6.         GISAID - CoVsurver mutations App [Internet]. [cited 2021 Jan 30]. Available from: https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/

7.         PANGO lineages [Internet]. [cited 2021 Jan 28]. Available from: https://cov-lineages.org/pangolin.html

8.         Rambaut A, Holmes EC, O’Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nature Microbiology [Internet]. 2020 Nov 1 [cited 2021 Jan 28];5(11):1403–7. Available from: https://doi.org/10.1038/s41564-020-0770-5

9.         Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.

10.        Tian Chen A, Altschuler K, Hei Zhan S, Alina Chan Y. COVID-19 CG: Tracking SARS-CoV-2 mutations by locations and dates of interest 1. [cited 2021 Jan 28]; Available from: https://doi.org/10.1101/2020.09.23.310565