**Supporting Information for**

**ViralVar: a web tool for multilevel visualization of SARS-CoV-2 genomes**

Arghavan Alisoltani1,2,3\*, Lukasz Jaroszewski4, Adam Godzik4, Arash Iranzadeh5, Lacy M. Simons2,3, Taylor Dean2,3, Ramon Lorenzo-Redondo2,3, Judd F. Hultquist2,3, Egon A. Ozer2,3\*

Affiliations

1Department of Microbiology-Immunology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

2Department of Medicine, Division of Infectious Diseases, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

3Center for Pathogen Genomics and Microbial Evolution, Havey Institute for Global Health, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

4University of California Riverside School of Medicine, Biosciences Division, Riverside, California, USA

5Department of Integrative Biomedical Sciences, Computational Biology Division, University of Cape Town, Cape Town, South Africa

\*Corresponding author contact: Egon Ozer ([e-ozer@northwestern.edu](mailto:e-ozer@northwestern.edu)) and Arghavan Alisoltani ([arghavan.alisoltanidehkordi@northwestern.edu](mailto:arghavan.alisoltanidehkordi@northwestern.edu))

Supplementary Figures



**Figure S1 |** Frequency of new cases and deaths in the USA. The data underlying the figures obtained from <https://data.cdc.gov/>.

****

**Figure S2 |** **Area plots reflecting the relative abundance of variants of concern and variants of interest collected in the USA over time.** **A)** Absolute frequency of indicated VOCs and VOIs over time in specimens collected between January 2020 and May 2022 in the USA (n = 1,739,737 sequences, from GISAID as of May 31, 2022). **B**) Absolute frequency of indicated VOCs and VOIs over time in specimens collected between November 2020 and March 2021. Specimens were divided into three age groups: children (up to 18 years), adults (18–64 years), and the elderly (65 years or more). The number of sequences per age group is indicated above each plot. Each subset of genomes was processed separately using the ViralVar ‘Lineage Dynamics’ module.



**Figure S3 |** Runtime performance for K-means clustering (including estimation of optimal number of clusters) and calculation of pairwise distance matrices for SARS-CoV-2 genomes (different sample size) and 86 Spike mutations. Each calculation repeated five times and average execution time reported (in seconds). All calculations were conducted in R on MacBook Pro (14-inch, 2021) with 8-core CPU, 14-core GPU, 16-core Neural Engine Accessory Kita and 32 GB memory. RML: reached memory limit.