

## Structural bioinformatics

# Coronavirus3D: 3D structural visualization of COVID-19 genomic divergence

Mayya Sedova, Lukasz Jaroszewski, Arghavan Alisoltani and Adam Godzik  \*

Division of Biomedical Sciences, University of California Riverside School of Medicine, Riverside, CA 92521, USA

\*To whom correspondence should be addressed.

Associate Editor: Arne Elofsson

Received on April 12, 2020; revised on May 18, 2020; editorial decision on May 22, 2020 ; accepted on May 25, 2020

## Abstract

**Motivation:** As the COVID-19 pandemic is spreading around the world, the SARS-CoV-2 virus is evolving with mutations that potentially change and fine-tune functions of the proteins coded in its genome.

**Results:** Coronavirus3D website integrates data on the SARS-CoV-2 virus mutations with information about 3D structures of its proteins, allowing users to visually analyze the mutations in their 3D context.

**Availability and implementation:** Coronavirus3D server is freely available at <https://coronavirus3d.org>.

**Contact:** adam.godzik@medsch.ucr.edu.

## 1 Introduction

The main challenge in the rapidly developing COVID-19 outbreak is the management of the current pandemic, but predicting its future course is quickly becoming a major focus. Differences in the societal responses, such as various levels of social distancing and screening/quarantine implementation are probably the main reason behind the different courses that COVID-19 takes in different countries and regions. But at the same time, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus is mutating, which might result in virus escape from diagnostic tests or virus resistance to therapeutic interventions. Over twenty-seven thousand SARS-CoV-2 genomes have been sequenced as of May 15, 2020 and their phylogenetic analysis identified the emergence of three major viral clades (GISAID as of May 15, 2020). Some of the widespread mutations observed in these clades result in amino-acid substitutions. Inspection of the corresponding protein structures strongly suggests that they may have an impact on the conformation and functions of the proteins they are found in and, possibly, on the COVID-19 outcomes. While there are no confirmed clinical differences between SARS-CoV-2 from different clades, the ongoing growth of the number of mutations create a high demand for the systematic analysis of non-synonymous mutations and their possible influence on the COVID-19 pandemics. This provided motivation for the development of the coronavirus3D server that provides a unique platform for exploring the distribution of the mutations in the context of the 3D structure of the proteins they are found in.

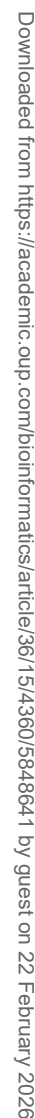
Information on the growing genetic diversity of SARS-CoV-2 is being studied intensively and continuously updated data can be obtained from resources such as GISAID (<https://www.gisaid.org>) or Nextstrain (<https://nextstrain.org>). At the same time, with the exception of the spike protein mutations, there are no

publicly available resources that provide analysis for all the other structurally characterized regions of the SARS-CoV-2 proteins.

## 2 Methods and server description

Coronavirus3D server integrates information about the three-dimensional structures of SARS-CoV-2 virus proteins from the PDB (<http://rcsb.org>) (Berman, 2000), with the data on SARS-CoV-2 genomic variations retrieved from China National Center for Bioinformation (CNCB) (<https://bigd.big.ac.cn/ncov?lang=en>). The server is updated automatically as new data becomes available, the date and details of the last update are listed on the top of the genome viewer panel. The Coronavirus3D website was developed with the Protal package (Sedova *et al.*, 2016) and 3D visualizations use the 3dmol.js library (Rego and Koes, 2015). The structural models of SARS-CoV-2 proteins without experimental structures were built using MODELLER (Webb and Sali, 2016) based on FFAS (Xu *et al.*, 2014) alignments.

The central page of the coronavirus server (see Fig. 1a,) provides an interactive view of the SARS-CoV-2 genome (GenBank ID: MN908947.3), with information on boundaries of the predicted proteins, currently available SARS-CoV-2 structures and a histogram of the amino-acid mutation frequency. If no SARS-CoV-2 structure is available, links are provided to the models based on the SARS-CoV structures. In the future we plan to incorporate *ab initio* models. Currently, we provide references to the resources for such predictions on the Help page. Using buttons on the top of the viewer or selecting specific regions with a mouse, users can zoom in to the display of the selected regions at higher resolution. Users can also select individual structures or models, which automatically displays information on the selected structure in the lower panels. Detailed



We acknowledge efforts of the teams at PDB, CNCB, GISAID and NextStrain for maintaining and distributing information on COVID-19 and all the individual laboratories that make their results public and available through these depositories.

## Funding

This work was supported in whole or in part by NIH institutes: NIAID under contract no. HHSN272201700060C and NIGMS by a grant GM118187.

*Conflict of Interest:* none declared.

## References

- Berman,H.M. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Hillen,H.S. et al. (2020) Structure of replicating SARS-CoV-2 polymerase. *PDB*.
- Laha,S. et al. (2020) Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *bioRxiv*.
- Pachetti,M. et al. (2020) Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.*, **18**, 179.
- Rego,N. and Koes,D. (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics*, **31**, 1322–1324.
- Rosas-Lemus,M. et al. (2020) The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv*.
- Sedova,M. et al. (2016) Protal: protein data visualization library for the web. *Bioinformatics*, **32**, 602–604.
- Wang,Q. et al. (2016) Immunodominant SARS coronavirus epitopes in humans elicited both enhancing and neutralizing effects on infection in non-human primates. *ACS Infect. Dis.*, **2**, 361–376.
- Webb,B. and Sali,A. (2016) Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.*, **86**, 9 1–29. 37.
- Xu,D. et al. (2014) FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics*, **30**, 660–667.