Bioinformatics, 36(15), 2020, 4360–4362
doi: 10.1093/bioinformatics/btaa550
Advance Access Publication Date: 29 May 2020
Applications Note

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 Coronaviusr3D website was developed with the Protael 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

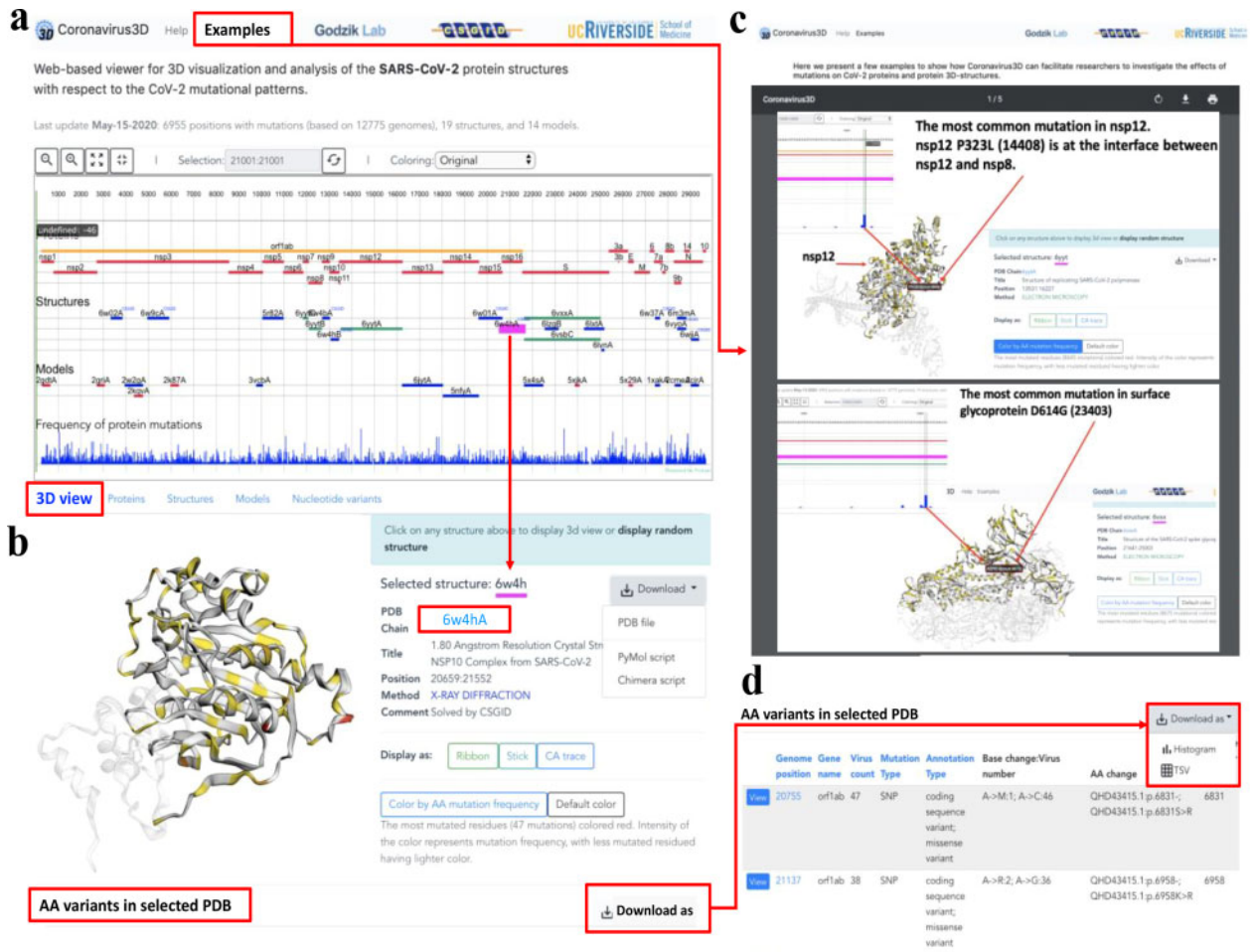

Fig. 1. The overview of the Coronavirus3D server. (**a**) Top panel of the main page contains a zoomable genome viewer. (**b**) The lower panel of the main page shows an interactive visualization of the structure selected above (6w4h). A pale yellow to reddish color spectrum in the 3D view of the protein structure represents low to high rates of mutations (counts of viral genomes with this mutation). (**c**) The 'Examples' tab provides examples of how Coronavirus3D can provide insights into the effects of mutations on structures of SARS-CoV-2 proteins. (**d**) A downloadable list of amino-acid mutations in the selected PDB structure

information about the functions of the user interface is provided in the help pages available via the link located at the top of the page.

The first of the lower level panels (Fig. 1b) provides interactive visualization of the selected structure or model, with an option for coloring the chain according to the mutation frequency. The example in Figure 1b shows the SARS-CoV-2 structure of the complex between nsp10 and nsp16 [PDB ID: 6w4h (Rosas-Lemus *et al.*, 2020)]. Because chain A was selected for viewing, this chain is shown in color, with the second chain shown with lower intensity. As seen in the figure, some mutations fall on the nsp10/nsp16 interface, possibly changing the stability of the complex. The second of the lower panels (Fig. 1d) provides a list of mutations in the selected protein (or in the selected genomic regions) that can be downloaded for further analysis.

3 Results

The Coronavirus3D server was designed to provide users with information and tools to carry out their own analysis of how mutations in the SARS-CoV-2 proteins may affect their 3D-structures and their functions. We show here two examples of such analyses. The first example is the most common mutation in RNA-directed RNA polymerase (RdRp or nsp12). This mutation, P323>L (genomic position 14 408), is located at the interface between nsp12 and nsp8 proteins in the RdRp complex, as shown on the experimental structure of the nsp7/nsp8/nsp12 complex [PDB ID : 6yyt (Hillen *et al.*, 2020)] (Fig. 1c, top). Mutations at this interface may change the strength of the interactions in the complex and its activation profile. Interestingly, genomes with this mutation were demonstrated to have significantly (∼3 times) higher mutation frequency as compared to genomes without this mutation (Pachetti *et al.*, 2020), which could be related to the overactivation of the RNA polymerase complex. The second example shows the most widespread mutation in Spike glycoprotein - D614>G (genomic position 23 403), visualized here on the experimental structure of the spike protein and the human ACE2 receptor [PDB ID : 6vxx (Laha *et al.*, 2020)] (Fig. 1c, bottom). This is the defining mutation of the G clade of the SARS-Cov-2. The corresponding position in SARS-CoV is part of an immunodominant epitope (Wang *et al.*, 2016). Interestingly, the mutations mentioned in these two examples are observed in practically the same set of genomes corresponding to the G clade. It is still unclear which of these mutations (if any) contributes to the apparent recent expansion of the G clade.

Acknowledgements

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) Protael: 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.