Bioinformatics, 38(18), 2022, 4418–4419
https://doi.org/10.1093/bioinformatics/btac517
Advance Access Publication Date: 21 July 2022
Applications Note

OXFORD

Sequence analysis

SpikePro: a webserver to predict the fitness of SARS-CoV-2 variants

Gabriel Cia[1], Jean Marc Kwasigroch[1], Marianne Rooman[1,2,*,†] and Fabrizio Pucci [1,2,†]

[1]Computational Biology and Bioinformatics, Université Libre de Bruxelles, Avenue F.D. Roosevelt 50, Brussels 1050, Belgium and [2]Interuniversity Institute of Bioinformatics in Brussels, Boulevard du Triomphe, Brussels 1050, Belgium

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the last two authors should be regarded as Joint Authors.
Associate Editor: Pier Luigi Martelli

Received on April 17, 2022; revised on July 7, 2022; editorial decision on July 14, 2022; accepted on July 20, 2022

Abstract

Motivation: The SARS-CoV-2 virus has shown a remarkable ability to evolve and spread across the globe through successive waves of variants since the original Wuhan lineage. Despite all the efforts of the last 2 years, the early and accurate prediction of variant severity is still a challenging issue which needs to be addressed to help, for example, the decision of activating COVID-19 plans long before the peak of new waves. Upstream preparation would indeed make it possible to avoid the overflow of health systems and limit the most severe cases.
Results: We recently developed SpikePro, a structure-based computational model capable of quickly and accurately predicting the viral fitness of a variant from its spike protein sequence. It is based on the impact of mutations on the stability of the spike protein as well as on its binding affinity for the angiotensin-converting enzyme 2 (ACE2) and for a set of neutralizing antibodies. It yields a precise indication of the virus transmissibility, infectivity, immune escape and basic reproduction rate. We present here an updated version of the model that is now available on an easy-to-use webserver, and illustrate its power in a retrospective study of fitness evolution and reproduction rate of the main viral lineages. SpikePro is thus expected to be great help to assess the fitness of newly emerging SARS-CoV-2 variants in genomic surveillance and viral evolution programs.
Availability and implementation: SpikePro webserver http://babylone.ulb.ac.be/SpikePro/.
Contact: marianne.rooman@ulb.be
Supplementary information: Supplementary data are available at *Bioinformatics* online.

Genomic surveillance has played a major role in combating the different waves of SARS-CoV-2 variants as it has enabled their early identification, characterization and mitigation (Oude Munnink *et al.*, 2021). Despite the weakening of the pandemic currently observed due to milder infections of the Omicron lineage, the appearance of antigenically highly divergent variants, which would lead to host immune escape and to new lineages of increased severity, is still possible (Markov *et al.*, 2022). It is therefore still extremely important to maintain and develop all genomic surveillance efforts.

Much computational and experimental research has been devoted to studying the impact of SARS-CoV-2 variants on transmissibility, the ability to escape from the host immune system and to become dominant at the level of viral population (Greaney *et al.*, 2021b; Starr *et al.*, 2020). However, it is still a challenge to make such predictions in an accurate and fast manner (Oude Munnink *et al.*, 2021). Our recently published computational model called SpikePro (Pucci and Rooman, 2021) meets this objective. It predicts the fitness of SARS-CoV-2 variants on the basis of the spike protein sequence and structure. Here, we present an updated version of the model along with a userfriendly webserver which makes SpikePro accessible to non-specialists. We describe below the main functionalities of the SpikePro webserver (Fig. 1) and refer the reader to Supplementary Material and our original publication (Cia *et al.*, 2022) for details.

• **Molecular-level insights of SARS-CoV-2 fitness and evolution.** SpikePro predicts the overall fitness Φ_i of a SARS-CoV-2 variant i with respect to the Wuhan lineage in terms of three fitness components ϕ_i^S, ϕ_i^{ACE2} and ϕ_i^{nAb} as:

$$\Phi_i = \phi_i^S \times \phi_i^{ACE2} \times \phi_i^{nAb}$$

where:

1. ϕ_i^S is defined as a function of the change in folding free energy ($\Delta\Delta G_S$) between the spike protein variant i and the Wuhan

© The Author(s) 2022. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/bioinformatics/article/38/18/4418/6647899 by Universite Libre de Bruxelles user on 23 January 2023

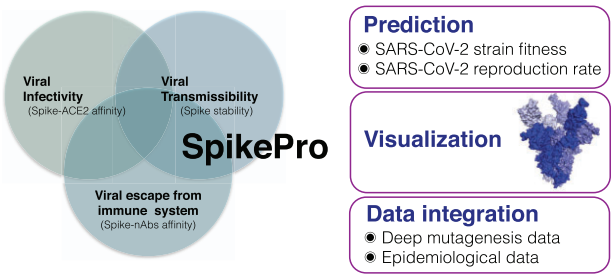

Fig. 1. Schematic overview of the main functionalities of the SpikePro webserver

lineage. This component captures the impact of mutations on the spike protein stability which is related to viral transmissibility;

2. ϕ_i^{ACE2} is a function of the change in binding free energy ($\Delta\Delta G_B$) between the complexes formed by the variant and Wuhan spike protein with ACE2. As this receptor is known to be the SARS-CoV-2 entry point into the host cells, this component predicts viral infectivity; and
3. ϕ_i^{nAb} is a function of the average change in binding affinity ($\Delta\Delta G_B$) between the variant and Wuhan spike protein for a set of neutralizing antibodies (nAbs). This component captures the impact of mutations on the escape of the virus from the host's humoral immune response.

The $\Delta\Delta G_S$ and $\Delta\Delta G_B$ values were estimated using our structure-based in-house predictors PoPMuSiC (Dehouck *et al.*, 2009) and BeAtMuSiC (Dehouck *et al.*, 2013), respectively.

The decomposition of the viral fitness into three components representing transmissibility, infectivity and viral immune escape helps users understand which molecular mechanism drives the overall fitness evolution of the SARS-CoV-2 variants with respect to the original Wuhan lineage.

• **SARS-CoV-2 basic reproduction rate prediction.** Based on the predicted overall fitness Φ, SpikePro estimates the basic reproduction rate of spike protein variants, R_0, which is a commonly monitored indicator for new emerging variants, describing the average number of secondary infections that each infected person causes in a naive population (Arora *et al.*, 2022).

• **Experimental and epidemiological data integration.** The webserver integrates experimental information about the effect of all single-site mutations in the spike protein on the viral escape and on the change in binding affinity for ACE2 taken from high-throughput mutagenesis experiments (Greaney *et al.*, 2021a,b; Starr *et al.*, 2020). In addition, it provides the frequency of occurrence of each mutation in global initiative on sharing all influenza data (GISAID) (Shu and McCauley, 2017), a weekly updated dataset of SARS-CoV-2 variant sequences observed worldwide.

• **Structure visualization.** The webserver includes a tool to visualize the three-dimensional structure of the spike protein trimer in different closed and open conformations, allowing straightforward analysis of the localization and spatial environment of the mutations.

In summary, the SpikePro webserver provides a series of efficient predictions of SARS-CoV-2 variant fitness as well as additional data and analysis tools aimed at providing molecular-level understanding of viral evolution. It therefore has a role to play as an additional surveillance tool for emerging variants.

To illustrate the power of the SpikePro method, we predicted and analyzed the basic reproduction rate R_0 of the main SARS-CoV-2 lineages and the evolution of their fitness components in a retrospective study from the start of the pandemic until now; the results are given in Supplementary Material. As an example, we show in Figure 2, the predicted evolution of SARS-CoV-2's R_0 as a function of time, averaged over all spike protein sequences of the GISAID database (Shu and McCauley, 2017). We observe its constant increase due to natural selection acting on the viral spike protein sequence.

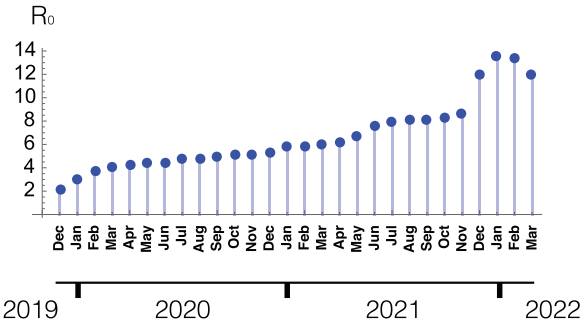

Fig. 2. SARS-CoV-2 R_0 predicted by SpikePro, averaged over all viral sequences collected from the GISAID database (Shu and McCauley, 2017), as a function of time. Details are given in Supplementary Material

Funding

We acknowledge financial support from the FNRS—Fund for Scientific Research through a COVID—Exceptional Research Project. F.P. and M.R. are FNRS postdoctoral researcher and research director, respectively, and G.C. benefits from a FNRS-FRIA PhD grant.

Conflict of Interest: none declared.

Data availability

The data underlying this article are available in the repository github.com/3BioCompBio/SpikeProSARS-CoV-2.

References

Arora,P. *et al.* (2022) No evidence for increased cell entry or antibody evasion by Delta sublineage AY.4.2. *Cell. Mol. Immunol.*, **19**, 449–452.

Cia,G. *et al.* (2022) Analysis of the neutralizing activity of antibodies targeting open or closed SARS-CoV-2 spike protein conformations. *IJMS*, **23**, 2078.

Dehouck,Y. *et al.* (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: poPMuSiC-2.0. *Bioinformatics*, **25**, 2537–2543.

Dehouck,Y. *et al.* (2013) BeAtMuSiC: prediction of changes in protein–protein binding affinity on mutations. *Nucleic Acids Res.*, **41**, W333–W339.

Greaney,A.J. *et al.* (2021a) Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection. *Sci. Transl. Med.*, **13**, eabi9915.

Greaney,A.J. *et al.* (2021b) Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe*, **29**, 463–476.

Markov,P.V. *et al.* (2022) Antigenic evolution will lead to new SARS-CoV-2 variants with unpredictable severity. *Nat. Rev. Microbiol.*, **20**, 251–252.

Oude Munnink,B.B. *et al.* (2021) The next phase of SARS-CoV-2 surveillance: real-time molecular epidemiology. *Nat. Med.*, **27**, 1518–1524.

Pucci,F. and Rooman,M. (2021) Prediction and evolution of the molecular fitness of SARS-CoV-2 variants: introducing SpikePro. *Viruses*, **13**, 935.

Shu,Y. and McCauley,J. (2017) GISAID: global initiative on sharing all influenza data—from vision to reality. *Euro Surveill.*, **22**, 30494.

Starr,T.N. *et al.* (2020) Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*, **182**, 1295–1310.

Downloaded from https://academic.oup.com/bioinformatics/article/38/18/4418/6647899 by Universite Libre de Bruxelles user on 23 January 2023