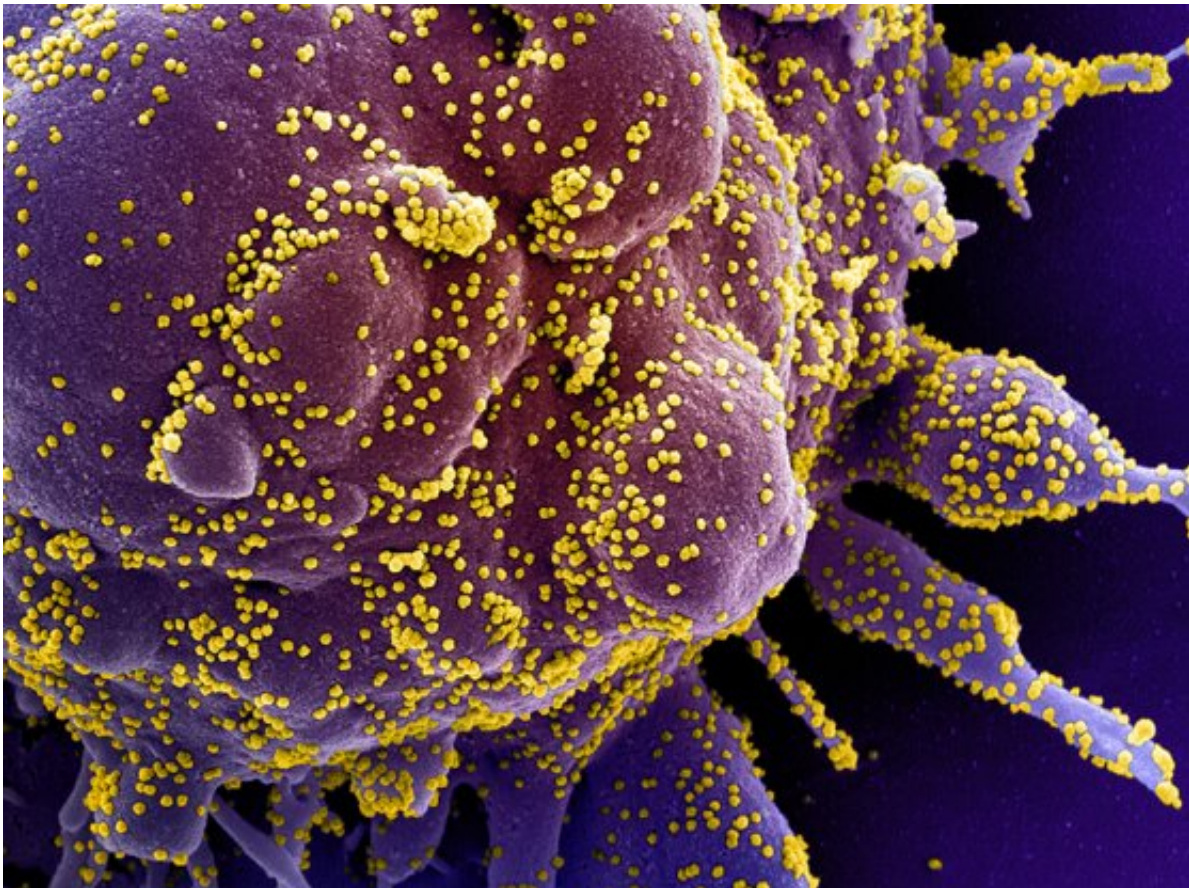


Computer model predicts dominant SARS-CoV-2 variants

May 25 2022, by Alessandra Dicorato



Colorized electron microscopy image of a cell (purple) infected with SARS-CoV-2 (yellow), isolated from a patient sample. Credit: National Institute of Allergy and Infectious Diseases, NIH

Scientists at the Broad Institute of MIT and Harvard and the University

of Massachusetts Medical School have developed a machine-learning model that can analyze millions of SARS-CoV-2 genomes and predict which viral variants will likely dominate and cause surges in COVID-19 cases. The model, called PyR0 (pronounced "pie-are-nought"), could help researchers identify which parts of the viral genome will be less likely to mutate and hence be good targets for vaccines that will work against future variants. The findings appear today in *Science*.

The researchers trained the [machine-learning model](#) using 6 million SARS-CoV-2 genomes that were in the GISAID database in January 2022. They showed how their tool can also estimate the effect of genetic [mutations](#) on the virus's fitness—its ability to multiply and spread through a population. When the team tested their model on viral genomic data from January 2022, it predicted the rise of the BA.2 variant, which became dominant in many countries in March 2022. PyR0 would have also identified the alpha variant (B.1.1.7) by late November 2020, a month before the World Health Organization listed it as a variant of concern.

The research team includes first author Fritz Obermeyer, a machine-learning fellow at the Broad Institute when the study began, and senior authors Jacob Lemieux, an instructor of medicine at Harvard Medical School and Massachusetts General Hospital, and Pardis Sabeti, an institute member at Broad, a professor at the Center for Systems Biology and the Department of Organismic and Evolutionary Biology at Harvard University, and a professor in the Department of Immunology and Infectious Disease at the Harvard T. H. Chan School of Public Health. Sabeti is also a Howard Hughes Medical Institute investigator.

PyR0 is based on a machine-learning framework called Pyro, which was originally developed by a team at Uber AI Labs. In 2020, three members of that team including Obermeyer and Martin Jankowiak, the study's second author, joined the Broad Institute and began applying the

framework to biology.

"This work was the result of biologists and geneticists coming together with [software engineers](#) and computer scientists," Lemieux said. "We were able to tackle some really challenging questions in public health that no single disciplinary approach could have answered on its own."

"This kind of machine learning-based approach that looks at all the data and combines that into a single prediction is extremely valuable," said Sabeti. "It gives us a leg up on identifying what's emerging and could be a potential threat."

The future of SARS-CoV-2

Researchers around the world have been working to predict the fitness of different SARS-CoV-2 viral variants since early in the pandemic. But previous models could not compare all variants simultaneously, or took days to process only a few thousand genomes.

By contrast, PyR0 can analyze millions of genomes—all of the publicly available SARS-CoV-2 data—in about an hour. It does this by grouping similar sequences together, and then defining "clusters" of genomes by the constellation of mutations they share. By focusing on mutations, which can appear in multiple variants, PyR0 has more statistical power than models that focus on viral variants.

Next, the model determines which mutations are becoming more common and estimates how quickly each mutation can cause the virus to spread. It also estimates how rapidly the number of cases of different variants will increase based on their genetic makeup.

By identifying which mutations are important for the fitness of particular variants, the model also offers biological insight into how

COVID-19 spreads and develops. For example, knowing the critical mutations can help scientists predict whether new variants will be more contagious or evade neutralizing antibodies, and can also help them decide which mutations to study in greater detail.

"The SARS-CoV-2 genome now has accumulated many mutations, so it becomes extremely challenging to interrogate all combinations of mutations," said Jankowiak, a machine-learning fellow at the Broad.

"The advantage of this kind of analysis is that it looks at the entire genome holistically, and may point to mutations or variants that are receiving less attention in the lab."

Early warning

The researchers say their study suggests that current increases in viral fitness stem from the virus's ability to escape immune responses. They add that [public health](#) officials, with advanced warning of a variant's sequence and characteristics, could implement specific measures to manage case counts. And knowing which mutations are contributing to a [variant](#)'s survival—and are thus not likely to change—can help researchers pick better targets for future vaccines.

New versions of this or similar models could further improve predictions by taking into account interactions between mutations. The researchers say that with further work, their model could help monitor other viruses that have enough genetic data.

"The amount of data that we have, together with the methods that we've developed, allow us to get a real-time view of the virus evolving in different locations around the world in a way that was just not possible during previous epidemics," said Obermeyer. "In 1917, people only knew if they had the flu, or they didn't. Now, we have a very precise view of thousands of different SARS-CoV-2 sub-lineages. That's just

amazing."

More information: Fritz Obermeyer et al, Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness, *Science* (2022). [DOI: 10.1126/science.abm1208](https://doi.org/10.1126/science.abm1208)

Provided by Broad Institute of MIT and Harvard

Citation: Computer model predicts dominant SARS-CoV-2 variants (2022, May 25) retrieved 6 September 2023 from <https://medicalxpress.com/news/2022-05-dominant-sars-cov-variants.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.