

Researchers develop rapid computer software to track pandemics as they happen

November 16 2021



Credit: Pixabay/CC0 Public Domain

Researchers at Georgia State University have created lightning-fast computer software that can help nations track and analyze pandemics, like the one caused by COVID-19, before they spread like wildfire around the globe.

The group of computer science and mathematics researchers says its new software is several orders of magnitude faster than existing [computer programs](#) and can process more than 200,000 novel virus genomes in less than two hours. The software then builds a clear visual tree of the strains and where they are spreading. This provides information that can be invaluable for countries making early decisions about lockdowns, quarantines, social distancing and testing during infectious disease outbreaks.

"The future of infectious outbreaks will no doubt be heavily data driven," said Alexander Zelikovsky, a Georgia State computer science professor who worked on the project.

The new software was co-created with Pavel Skums, assistant professor of computer science, Mark Grinshpon, principal senior lecturer of mathematics and statistics, Daniel Novikov, a computer science Ph.D. student, and two former Georgia State Ph.D. students—Sergey Knyazev (now a postdoctoral scholar at the University of California at Los Angeles) and Pelin Icer (now a postdoctoral scholar at Swiss Federal Institute of Technology, ETH Zürich).

Their paper describing the new approach, Scalable Reconstruction of SARS-CoV-2 Phylogeny with Recurrent Mutations, was published in the *Journal of Computational Biology*.

"The COVID-19 pandemic has been an unprecedented challenge and opportunity for scientists," said Skums, who noted that never before have researchers around the world sequenced so many complete genomes of any virus. The strains of SARS-CoV-2 are uploaded onto the free global GISAID database, where they can be data-mined and studied by any scientist. Zelikovsky, Skums and their colleagues analyzed more than 300,000 different GISAID strains for their new work.

"There are over 5 million genomes in the GISAID database now," said Zelikovsky. "Scientists around the globe are probably sequencing a new variant almost every hour."

Zelikovsky said that this astounding amount of data allows scientists to see the evolution of the virus in action in real time—if we have software capable of rapidly analyzing it.

In the early days of the pandemic, in March 2020, scientists were working much more slowly. Scientists thought the virus had first arrived on our shores in the state of Washington in February. However, later sequencing [presented in a paper](#) by Skums and his colleagues showed the arcs of viral variants traveling across countries and oceans. With new studies, scientists learned that the virus had also likely arrived quietly in New York City in February, from strains originating in Europe.

Back then, scientists were sequencing data too slowly to capture the true migration of this global virus and its mutations in real time.

"The programs were not fast enough, not scalable enough," said Skums. "The algorithms were not equipped to handle huge amounts of data." It could take hours or days to process even a small subset of viral genomes, he said.

Zelikovsky, Skums and their colleagues created a novel [algorithm](#) for viral sequencing called SPHERE (Scalable PHylogEny with Recurrent mutations.) SPHERE can rapidly handle huge amounts of real-time data and create evolutionary trees of the virus and its mutations. These visualizations can be easily grasped at a glance. The computer program itself is freely available for download to any researcher in the world.

When the researchers applied their algorithm to genomes from the GISAID database, they found their SPHERE approach to be highly

reliable in tracking the way the virus was spreading. SPHERE can help scientists explore how a [virus](#) is evolving in real time.

"We can see how the mutations spread from country to country and region to region," said Zelikovsky. "We can determine how lockdowns and closures impact spread. This has consequences for government policy."

The SPHERE algorithm could prove invaluable in future pandemics.

"You could track down chains of transmission very quickly," said Zelikovsky. Seeing those chains will help governments to make sound decisions about social policies such as distancing or lockdowns during times of high transmission.

SPHERE can also show the impact of different approaches to outbreaks. For instance, said Skums, Sweden took a more relaxed approach to the COVID-19 pandemic than other Nordic countries. An analysis of the sequencing data shows that Swedes have longer "transmission chains." This means that in Sweden, one strain is able to infect many more people, one by one.

"The danger of long chains is that a new strain may appear," said Zelikovsky. "And one of those strains may be a variant that is very good at infecting people."

These kinds of insights will help us, should we face another global pandemic.

"The tools we and others have developed can be used anywhere for any [outbreak](#)," said Zelikovsky. "That is the beauty of computer science."

More information: Daniel Novikov et al, Scalable Reconstruction of

SARS-CoV-2 Phylogeny with Recurrent Mutations, *Journal of Computational Biology* (2021). [DOI: 10.1089/cmb.2021.0306](https://doi.org/10.1089/cmb.2021.0306)

Provided by Georgia State University

Citation: Researchers develop rapid computer software to track pandemics as they happen (2021, November 16) retrieved 19 April 2024 from <https://phys.org/news/2021-11-rapid-software-track-pandemics.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.