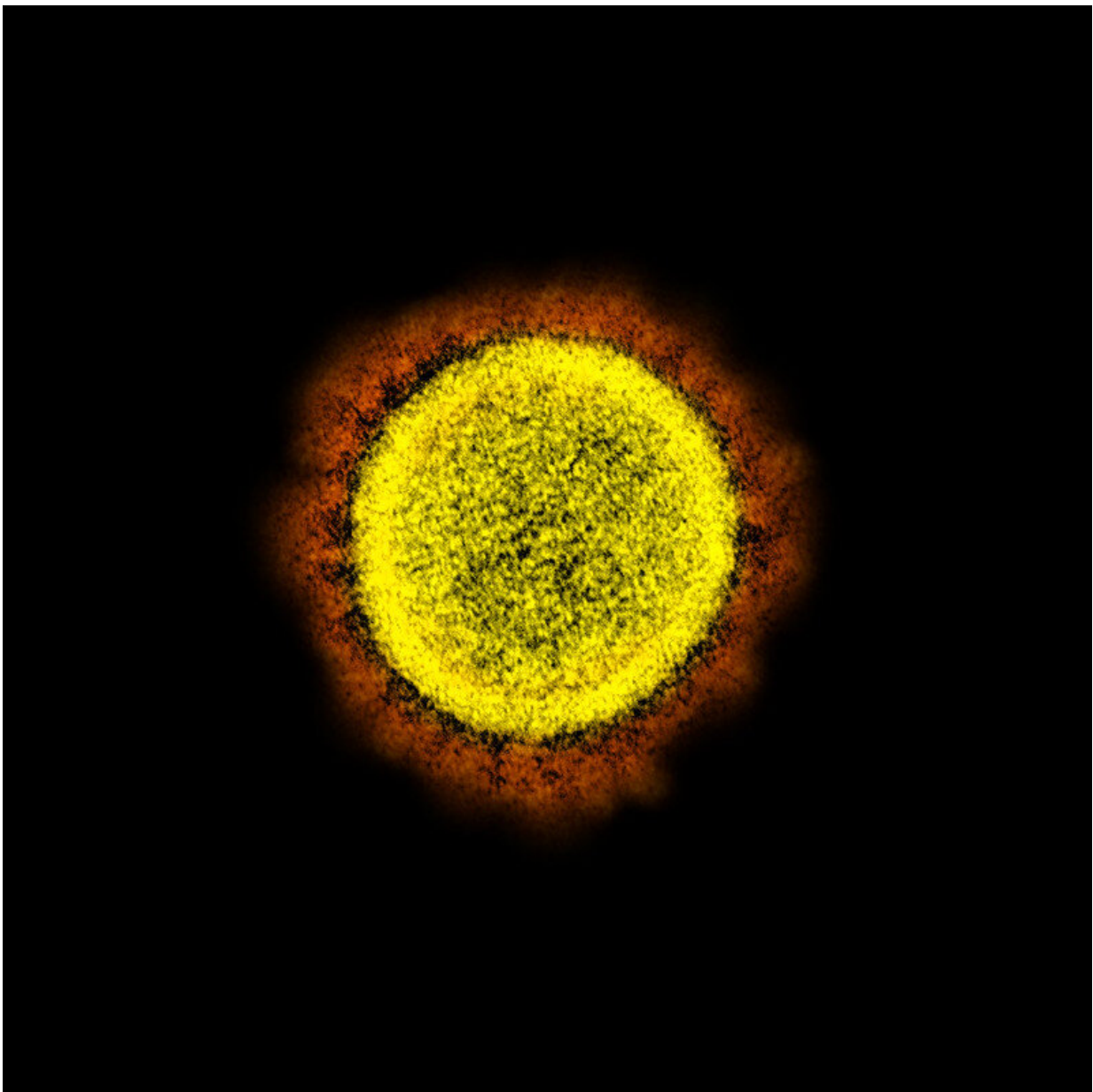


Origins of an outbreak: A recombination detection algorithm to find the source of SARS-CoV-2

March 31 2021, by David Pacchioli



Novel Coronavirus SARS-CoV-2 Transmission electron micrograph of SARS-CoV-2 virus particles, isolated from a patient. Image captured and color-enhanced at the NIAID Integrated Research Facility (IRF) in Fort Detrick, Maryland. Credit: National Institute of Allergy and Infectious Diseases, NIH

It was late January 2020 when Maciej Boni realized that the COVID-19 pandemic was about to take over his life.

Boni, associate professor of biology, is an epidemiologist with extensive expertise in viral evolution, including a recent focus on human and avian flu. When COVID-19 hit, he tapped into a network of colleagues around the world, quickly joining an international team intent on tracking the outbreak to its origins.

Coronaviruses like SARS-CoV-2, Boni knew, are highly recombinant, each a genetic mash-up of bits and pieces picked up and discarded through generations of evolution. As a [graduate student](#), he had created the recombination detection algorithm 3SEQ, the most accurate method yet devised for identifying recombinant viruses, and his research group in Penn State's Center for Infectious Disease Dynamics continues to maintain this important tool.

"So I thought, why not see how highly recombinant SARS-CoV-2 is?" he said.

The first reason for wanting to know the viral origins of an outbreak is to stop it. "Identify the point source and close a poultry market, close a wet market, isolate a single district before it's gotten to thousands of people," as Boni said. In the case of SARS-CoV-2, however, the outbreak had

already spread too far for that kind of intervention. If he and other experts could determine where the [virus](#) had come from, they would have a better chance at predicting where it was going. Understanding the [evolutionary history](#), moreover, would be critical for preventing future outbreaks.

To untangle the details of the SARS-CoV-2 genome, Boni and his colleagues used bioinformatics to pull out the recombinant segments.

"[That] left us with two or three major segments that, as far as we can tell, have not been broken up and pasted back together," he said. Using these fixed elements as a kind of evolutionary backbone, they created a family tree of all the coronaviruses they could identify in what was left. Within that panoply, they calculated that SARS-CoV-2 and its closest relative, a bat virus called RaTG13, diverged from a common ancestor between 40 and 70 years ago.

That means SARS-CoV-2 has been circulating in bats for decades, Boni says. What's more, one of the older traits that SARS-CoV-2 shares with RaTG13 and other close relatives is its receptor-binding site, the genetic mechanism that enables the virus to recognize and bind to receptors inside the human lung.

"The receptor binding site was not acquired by recombination from another virus," he explained. "That's something that just exists in bats—and in pangolins, it turns out. It's just a trait of these specific bat coronaviruses that they can also infect humans."

The scary part? "There are probably dozens or hundreds of other viruses on this viral lineage, some of which are ready to jump to humans whenever there's an opportunity," Boni says. The key to preventing the next outbreak, then, is preventing those opportunities, along with improved screening so that where crossover does occur, any further

spreading can be quickly minimized."

By late February 2020, however, the present outbreak was commanding Boni's full attention, as the full scope of the threat became clear. "I quickly wrapped up all the evolutionary work and started shifting to epidemiology," he says.

For this work, he has teamed up with Ephraim Hanks, associate professor of statistics, and Justin Pritchard, assistant professor of bioengineering, to assist and advise the state departments of health for Pennsylvania, Massachusetts and Rhode Island. Using data provided by each state, he explained, the three are conducting statistical analyses to help hospitals forecast future needs.

Long-term forecasting, he stressed, is next to impossible, because there is too large an unknown variable: human behavior. But what they can do with existing data is provide more accurate estimates that can help health officials get a better handle on the present state of the epidemic.

By feeding thousands of data points into their mathematical model, they get estimates of factors such as the percentage of 40- to 49-year-olds infected with the virus who become hospitalized, the percentage of hospitalized patients that are moved into the intensive care unit (ICU), and the average length of stay in the ICU or length of time on a ventilator.

With enough data from reported cases and a finely tuned model, they can then begin to get a better handle on the number of people who were infected, but did not report that they were.

"The most valuable thing we can do," Boni said, "is to provide states with [a more accurate estimate of] the total number of people who have been infected."

More information: 3SEQ Recombination Detection Algorithm:
mol.ax/software/3seq/

Provided by Pennsylvania State University

Citation: Origins of an outbreak: A recombination detection algorithm to find the source of SARS-CoV-2 (2021, March 31) retrieved 20 April 2024 from
<https://phys.org/news/2021-03-outbreak-recombination-algorithm-source-sars-cov-.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.