

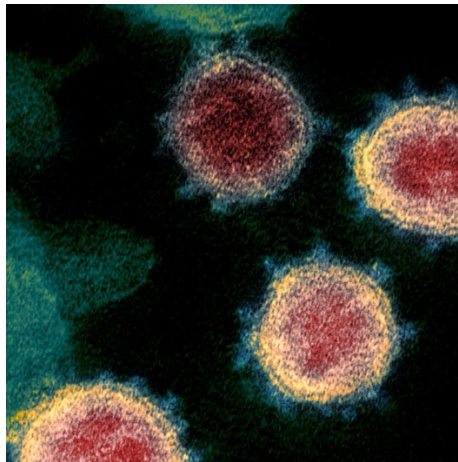
Media Inquiries:

Kendra Snyder
Department of Communications
212-496-3419
ksnyder@amnh.org
amnh.org

October 29, 2024

New Software Flags Emergence of COVID Variants

New surveillance technique could be used in future pandemics



Researchers have demonstrated a new technique for COVID surveillance that can signal the rise of new variants before they are widespread. The study, led by the American Museum of Natural History and Columbia University and published today in the journal [Genome Research](#), presents a way to track diversity across millions of genomes sequenced during the height of the COVID-19 pandemic using new surveillance software and points to its potential use in future pandemics.

During the COVID-19 pandemic, researchers all over the world used analyses of genomic data to guide policy. Sequences of SARS-CoV-2, the virus that causes COVID-19, revealed rapid evolution of new strains that were more contagious or more severe than the original virus.

“The architecture of the COVID-19 pandemic is marked by periods of transition, tipping a population towards an emerging variant of concern, followed by its near complete sweep to dominance,” said Apurva Narechania, lead author of the new study and a senior bioinformaticist in the Museum’s Institute for Comparative Genomics. “Speed is key to responding to these evolving strains, but the traditional way of analyzing these sequences slows down surveillance techniques.”

Working closely with colleagues in public health, Narechania looked to long-standing definitions of diversity that are used in ecology. In particular, the research team focused on Hill numbers, or the effective number of species in a sample, which provides a simple metric for comparing species diversity across environments. The more diverse the sample, the higher the Hill number. The new pandemic

surveillance software developed by the American Museum of Natural History-Columbia University team adopts this ecological approach, but in the place of species, the researchers use strings of sequence information, and in the place of environments, they use genomes.

The researchers tested this software on COVID-19 sequence data from multiple countries, including the United Kingdom, United States, and South Africa, finding that it accurately predicts the arc of variant emergence before the onset of sickness in the population.

“The viral surveillance we had during COVID-19 was a few weeks to months behind the edge of the pandemic curve,” said Barun Mathema, an associate professor of epidemiology at Columbia University’s Mailman School of Public Health and a corresponding author on the study. “In a crisis of COVID-19’s scale and speed, eliminating this analysis lag can mean the difference between timely, reasonable public health response and failure to understand and anticipate the disease’s next turn.”

The software, [now on GitHub](#), and freely available to non-commercial entities, cannot characterize new variants, but it can forewarn public health officials when a new strain is on the horizon and signal when more in-depth bioinformatics tools are needed. The researchers point to the software’s ability to detect new variants in wastewater as a particularly impactful potential application.

“We show that tracing a pandemic curve with these new metrics enables the use of sequence data as a real-time sensor, tracking both the emergence of variants over time and the extent of their spread,” Mathema said. “Our technique affords public health institutions the opportunity to create actionable policy based on a simple, quantitative measure.”

Narechania didn’t originally set out to solve this particular disease surveillance issue. The project began in 2020 with a focus on bacteria, which like viruses, can be very difficult to map onto an evolutionary tree.

“I don’t think I would have had this idea if I didn’t work at the Museum,” he said. “The original intent was to reimagine how we understand microbial species, but after COVID emerged, we turned our attention to tracing the pandemic’s natural history and testing whether it could be an effective tool in disease monitoring. The answer was a resounding yes.”

Support for this study was provided, in part, by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (grant # R01AI151173), the Centers for Disease Control (grant # 75D30121C11102/000HCVL1-2021-55232), and the Women’s Committee of the Children’s Hospital of Pennsylvania.

Other authors on the paper include Dean Bobo, American Museum of Natural History and Columbia University; Kevin Deitz and Rob Desalle, American Museum of Natural History; and Paul Planet, American Museum of Natural History, Children’s Hospital of Pennsylvania, and the University of Pennsylvania.

Study doi: 10.1101/gr.278594.123

ABOUT THE AMERICAN MUSEUM OF NATURAL HISTORY (AMNH)

The American Museum of Natural History, founded in 1869 with a dual mission of scientific research and science education, is one of the world’s preeminent scientific, educational, and cultural institutions. The Museum encompasses more than 40 permanent exhibition halls, galleries for temporary exhibitions, the Rose Center for Earth and Space including the Hayden Planetarium, and the Richard Gilder Center for Science, Education, and Innovation. The Museum’s scientists draw on a world-class permanent collection of more than 30 million specimens and artifacts, some of which are billions of years old, and on one of the largest natural history libraries in the world. Through its Richard Gilder Graduate School, the Museum

offers two of the only free-standing, degree-granting programs of their kind at any museum in the U.S.: the Ph.D. program in Comparative Biology and the Master of Arts in Teaching (MAT) Earth Science residency program. Visit amnh.org for more information.

#

Images: This transmission electron microscope image shows SARS-CoV-2, the virus that causes COVID-19—isolated from a patient in the U.S.
Courtesy of NIAID