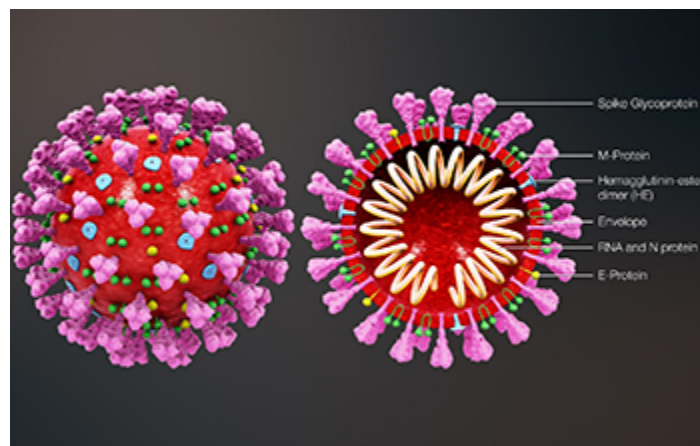




Research News

COVID-19 analysis performed with galaxy bioinformatics platform

Texas Advanced Computing Center provides large-scale infrastructure for analysis of thousands of genomes



NSF is supporting large-scale computer infrastructure for COVID-19 analyses.

[Credit and Larger Version \(/discoveries/disc_images.jsp?cntn_id=300525&org=NSF\)](https://discoveries.nsf.gov/disc-images.jsp?cntn_id=300525&org=NSF)

May 6, 2020

Scientists don't know how the SARS-CoV-2 virus, which causes COVID-19, will evolve.

About 100 organizations worldwide, mainly academic labs and genome sequencing facilities, have already contributed genomic data to the study of the pandemic. Genomic data is critical because it helps identify how the virus is evolving and, in turn, how it might be stopped.

"The community wasn't expecting this much data this quickly," said Sergei Pond, a biologist at Temple University in Philadelphia.

Pond and Anton Nekrutenko of Penn State are collaborating on the [Galaxy \(/cgi-bin/good-bye?https://covid19.galaxyproject.org/\)](https://covid19.galaxyproject.org/) project, one of the world's largest, most successful, web-based bioinformatics platforms.

"We have quite a user base, and have numerous instances across the world -- the biggest instance here in the U.S. is run out of the [Texas Advanced Computing Center \(/cgi-bin/good-bye?https://www.tacc.utexas.edu/-/covid-19-analysis-performed-with-galaxy-bioinformatics-platform\)](https://www.tacc.utexas.edu/-/covid-19-analysis-performed-with-galaxy-bioinformatics-platform)," said

Nekrutenko, a biochemist and molecular biologist.

Since 2013, TACC has powered data analyses for a large percentage of Galaxy users, allowing researchers to solve tough problems quickly and seamlessly in cases where their computers or campus clusters are not sufficient.

"Galaxy uses open source tools and public cyberinfrastructure for transparent, reproducible analyses of viral datasets -- it's free and promotes good practices," Nekrutenko said. "We run hundreds of thousands of analyses per month, and we're spiking now in terms of usage and viral analyses."

The researchers perform the majority of their analyses using parallel processing and big data analytics on TACC's [Stampede 2](https://www.nsf.gov/awardsearch/showAward?AWD_ID=1663578&HistoricalAwards=false) <https://www.nsf.gov/awardsearch/showAward?AWD_ID=1663578&HistoricalAwards=false> supercomputer, and the [Jetstream](https://www.nsf.gov/awardsearch/showAward?AWD_ID=1445604) <https://www.nsf.gov/awardsearch/showAward?AWD_ID=1445604> supercomputer led by Indiana with TACC as a partner.

"These important scientific findings underline the value of NSF's nearly 40-year investment in advanced cyberinfrastructure resources and services to enable national and international research collaborations that address critical problems," said Manish Parashar, director of NSF's Office of Advanced Cyberinfrastructure.

Galaxy employs the Bridges platform at the Pittsburgh Supercomputing Center for genome assembly jobs that require large amounts of shared memory. These systems are allocated through the [Extreme Science and Engineering Discovery Environment](https://www.xsede.org/) ([/cgi-bin/good-bye?https://www.xsede.org/](https://www.xsede.org/)), which awards supercomputer resources and expertise to researchers and is funded by the [National Science Foundation](https://www.nsf.gov/awardsearch/showAward?AWD_ID=1548562&HistoricalAwards=false) <https://www.nsf.gov/awardsearch/showAward?AWD_ID=1548562&HistoricalAwards=false>.