



Research News

National Science Foundation-funded scientists developing more accurate statistical models of COVID-19

Researchers mining public databases for genomic and associated coronavirus data



NSF-funded researchers are improving mathematical models used to predict coronavirus spread. [Credit and Larger Version \(/discoveries/disc_images.jsp?cntn_id=300548&org=NSF\)](/discoveries/disc_images.jsp?cntn_id=300548&org=NSF)

May 7, 2020

Armed with a rapid response research grant from the [National Science Foundation](https://www.nsf.gov/awardsearch/showAward?AWD_ID=2028986&HistoricalAwards=false) https://www.nsf.gov/awardsearch/showAward?AWD_ID=2028986&HistoricalAwards=false, [University of Oregon](https://around.uoregon.edu/content/uo-researcher-hopes-improve-covid-19-disease-models) ([/cgi-bin/good-bye?https://around.uoregon.edu/content/uo-researcher-hopes-improve-covid-19-disease-models](https://around.uoregon.edu/content/uo-researcher-hopes-improve-covid-19-disease-models)) biologist Stilianos Louca and his colleagues are mining public databases for genomic and associated data on the coronavirus that causes COVID-19.

Unlike the on-the-ground approach scientist John Snow used in the mid-1850s to find the source of a cholera outbreak in London, Louca is working on computers. His hope is to model a phylogenetic tree with predictive power to help guide medical decisions and public policies on the disease.

Phylogenetic trees constructed from viral genomes sampled from patients contain information about the historical pattern of transmission and dispersal of infectious diseases. Mathematical models of evolution allow researchers to infer critical epidemiological parameters, such as transmission rates, from information encoded in phylogenetic trees.

"Our goal is to develop more accurate statistical methods for estimating epidemiological parameters of infectious diseases from phylogenetic data, such as transmission rates and the basic reproduction ratio, and apply these methods to improving our understanding and predictions for COVID-19," said Louca.

Sequenced viral genomes are submitted in real-time from researchers around the world to two primary, open-access databases: GenBank of the National Center for Biotechnology Information; and the GISAID Initiative, originally known as the Global Initiative on Sharing All Influenza Data.

Genomes are usually submitted along with other data, including information on cities, countries and sampling dates, which provide valuable information for modeling the spread of the epidemic, Louca said.

In the new project, scientists are aiming to clarify the epidemiological insights that can be reliably inferred from phylogenetic trees and develop new approaches to characterize COVID-19 transmission. Additionally, Louca said, the researchers hope to determine which environmental, biological and policy factors affect the spread of COVID-19 based on the phylogenetic data.

"Many kinds of data are needed for accurate forecasting of this pandemic," said Sam Scheiner, a program director in NSF's Division of Environmental Biology. "This project adds past history to data based on current testing."

-- NSF Public Affairs, researchnews@nsf.gov (<mailto:researchnews@nsf.gov>)