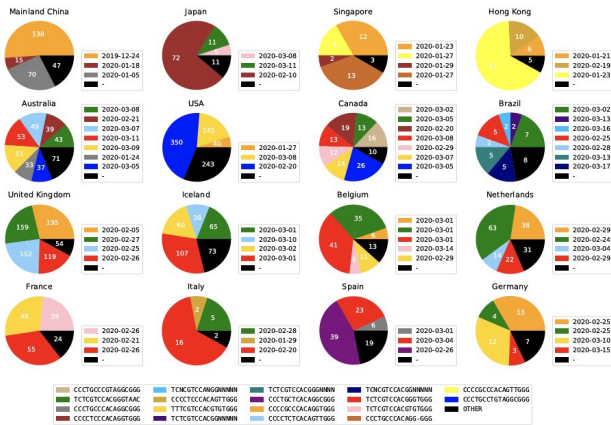


# Genetic tracing 'barcode' is rapidly revealing COVID-19's journey and evolution

17 April 2020, by Britt Faulstick



Major SARS-CoV-2 genetic subtypes in countries/regions with the most sequences (indicating date subtype was first sequenced in that country/region). Subtypes with less than 5% abundance are plotted as "OTHER." Credit: Drexel University

Drexel University researchers have reported a method to quickly identify and label mutated versions of the virus that causes COVID-19. Their preliminary analysis, using information from a global database of genetic information gleaned from coronavirus testing, suggests that there are at least six to 10 slightly different versions of the virus infecting people in America, some of which are either the same as, or have subsequently evolved from, strains directly from Asia, while others are the same as those found in Europe.

First developed as a way of parsing [genetic samples](#) to get a snapshot of the mix of bacteria, the genetic analysis tool teases out patterns from volumes of genetic information and can identify whether a virus has genetically changed. They can then use the pattern to categorize viruses with [small genetic differences](#) using tags called Informative Subtype Markers (ISM).

Applying the same method to process viral genetic data can quickly detect and categorize slight genetic variations in the SARS-CoV-2, the novel coronavirus that causes COVID-19, the group reported in a paper recently posted on the preliminary research archive, *bioRxiv*. The genetic analysis tool that generates these labels is publicly available for COVID-19 researchers on GitHub.

"The types of SARS-CoV-2 viruses that we see in tests from Asia and Europe is different than the types we're seeing in America," said Gail Rosen, Ph.D., a professor in Drexel's College of Engineering, who led the development of the tool. "Identifying the variations allows us to see how the virus has changed as it has traveled from population to population. It can also show us the areas where social distancing has been successful at isolating COVID-19."

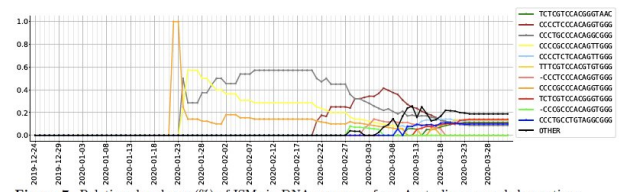


Figure 7. Relative abundance (%) of ISMs in DNA sequences from Australia as sampled over time.

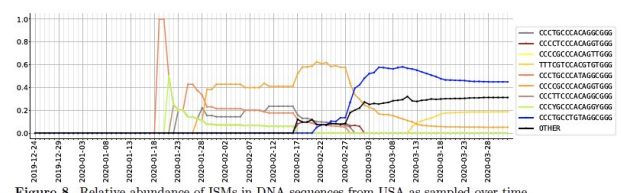
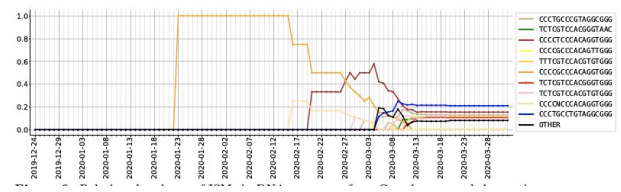


Figure 8. Relative abundance of ISMs in DNA sequences from USA as sampled over time.

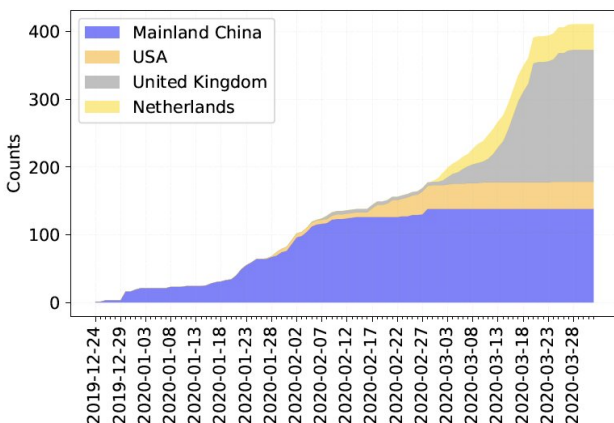


Relative abundance of ISMs in DNA sequences from Canada as sampled over time. Credit: Drexel University

The ISM tool, developed by Rosen and a focused team including doctoral student Zhengqiao Zhao and Bahrad A. Sokhansani, Ph.D., an independent researcher and intellectual property attorney, is particularly useful because it does not require analysis of the full genetic sequence of the virus to identify its mutations. In the case of SARS-CoV-2, this means reducing the 30,000-base-long genetic code of the virus to a subtype label 17 bases long.

"It's the equivalent of scanning a barcode instead of typing in the full product code number," Rosen said. "And right now, we're all trying to get through the grocery store a bit faster. For scientists this means being able to move to higher-level analysis much faster. For example, it can be a faster process in studying which virus versions could be affecting [health outcomes](#). Or, public health officials can track whether new cases are the result of local transmission or coming from other regions of the United States or parts of the world."

While these genetic differences might not be enough to delineate a new strain of virus, Rosen's group suggests understanding these genetically significant "subtypes," where they're being found and how prevalent they are in these areas is data granular enough to be useful.



Stacked plot of the number of sequences of the reference sequence ISM subtype (CCCCGCCACAGGTGGG).  
Credit: Drexel University

"This allows us to see the very specific fingerprint of COVID-19 from each region around the world, and to look closely at smaller regions to see how it is different," Rosen said. "Our preliminary analysis, using publicly available data from across the world, is showing that the combination of subtypes of virus found in New York is most similar to those found in Austria, France and Central Europe, but not Italy. And the subtype from Asia, that was detected here early in the pandemic has not spread very much, instead we are seeing a new subtype that only exists in America as the one most prevalent in Washington state and on the west coast."

In addition to helping scientists understand how the virus is changing and spreading, this method can also reveal the portion of its genetic code that appears to remain resistant to mutations—a discovery that could be exploited by treatments to combat the virus.

"We're seeing that the two parts of the [virus](#) that seem not to be mutating are the ones responsible for its entry into healthy cells and packaging its RNA," Rosen said. "Both of these are important targets for understanding the body's immune response, identifying antiviral therapeutics and designing vaccines."

**More information:** Zhengqiao Zhao et al. Characterizing geographical and temporal dynamics of novel coronavirus SARS-CoV-2 using informative subtype markers, *bioRxiv* (2020). [DOI: 10.1101/2020.04.07.030759](https://doi.org/10.1101/2020.04.07.030759)

Provided by Drexel University

APA citation: Genetic tracing 'barcode' is rapidly revealing COVID-19's journey and evolution (2020, April 17) retrieved 23 January 2021 from <https://phys.org/news/2020-04-genetic-barcode-rapidly-revealing-covid-.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.*