

Study identifies new 'hidden' gene in COVID-19 virus

November 10 2020



Credit: Pixabay/CC0 Public Domain

Researchers have discovered a new "hidden" gene in SARS-CoV-2—the virus that causes COVID-19—that may have contributed to its unique biology and pandemic potential. In a virus that only has about 15 genes

in total, knowing more about this and other overlapping genes—or "genes within genes"—could have a significant impact on how we combat the virus. The new gene is described today in the journal *eLife*.

"Overlapping [genes](#) may be one of an arsenal of ways in which coronaviruses have evolved to replicate efficiently, thwart host immunity, or get themselves transmitted," said lead author Chase Nelson, a postdoctoral researcher at Academia Sinica in Taiwan and a visiting scientist at the American Museum of Natural History. "Knowing that overlapping genes exist and how they function may reveal new avenues for [coronavirus](#) control, for example through antiviral drugs."

The research team identified ORF3d, a new overlapping gene in SARS-CoV-2 that has the potential to encode a protein that is longer than expected by chance alone. They found that this gene is also present in a previously discovered pangolin coronavirus, perhaps reflecting repeated loss or gain of this gene during the evolution of SARS-CoV-2 and related viruses. In addition, ORF3d has been independently identified and shown to elicit a strong antibody response in COVID-19 patients, demonstrating that the [new gene](#)'s protein is manufactured during human infection.

"We don't yet know its function or if there's clinical significance," Nelson said. "But we predict this gene is relatively unlikely to be detected by a T-cell response, in contrast to the antibody response. And maybe that has something to do with how the gene was able to arise."

At first glance, genes can seem like written language in that they are made of strings of letters (in RNA viruses, the nucleotides A, U, G, and C) that convey information. But while the units of language (words) are discrete and non-overlapping, genes can be overlapping and multifunctional, with information cryptically encoded depending on where you start "reading." Overlapping genes are hard to spot, and most

scientific computer programs are not designed to find them. However, they are common in viruses. This is partly because RNA viruses have a high mutation rate, so they tend to keep their gene count low to prevent a large number of mutations. As a result, viruses have evolved a sort of data compression system in which one letter in its genome can contribute to two or even three different genes.

"Missing overlapping genes puts us in peril of overlooking important aspects of viral biology," said Nelson. "In terms of genome size, SARS-CoV-2 and its relatives are among the longest RNA viruses that exist. They are thus perhaps more prone to 'genomic trickery' than other RNA viruses."

Prior to the pandemic, while working at the Museum as a Gerstner Scholar in Bioinformatics and Computational Biology, Nelson developed a computer program that screens genomes for patterns of genetic change that are unique to overlapping genes. For this study, Nelson teamed up with colleagues from institutions including the Technical University of Munich and the University of California, Berkeley, to apply this software and other methods to the wealth of new sequence data available for SARS-CoV-2. The group is hopeful that other scientists will investigate the gene they discovered in the lab to define its function and possibly determine what role it might have played in the emergence of the pandemic [virus](#).

More information: Chase W Nelson et al, Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic, *eLife* (2020). [DOI: 10.7554/eLife.59633](https://doi.org/10.7554/eLife.59633)

Provided by American Museum of Natural History

Citation: Study identifies new 'hidden' gene in COVID-19 virus (2020, November 10) retrieved 2 May 2024 from <https://phys.org/news/2020-11-hidden-gene-covid-virus.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.