
SARS-COV-2 VARIANTS OF CONCERN REMIND US THAT WE UNDERESTIMATE BIOLOGY AT OUR OWN PERIL

Samira Mubareka, Andrew McArthur and Sandrine Moreira | March 16, 2021

Samira Mubareka is a Clinician-Scientist at the University of Toronto and lead for the Sunnybrook Emerging and Respiratory Viruses Translational Research Program (SERV). She is also a contributor to the Ontario Provincial COVID-19 Genomics Network.

Andrew McArthur is the inaugural David Braley Chair in Computational Biology in McMaster University's Global Nexus for Pandemics and Biological Threats and contributes to genomic surveillance for SARS-CoV-2 Variants of Concern in Ontario.

Sandrine Moreira is program lead for Public Health Genomics and Bioinformatics at the Institut national de santé publique du Québec. She coordinates the genomic sequencing of SARS-COV-2 for the province of Québec and is also Associate Professor at Université de Montréal.

Imagine a tiny molecular machine armed with a program to infect human cells, subvert their defenses and make copies of itself to infect neighboring cells and spread to other humans. This strategy is used by SARS-CoV-2 among other viruses, and the code for this program is the virus's genome. Deciphering this code is important for detecting, tracking, and understanding SARS-CoV-2, whereby changes at the most fundamental and molecular level may have implications for individuals, populations and economies around the world.

Coronaviruses are highly proficient at adapting and are thus able to infect a diversity of hosts, from camels to humans. In some instances, such as with SARS-CoV-2, this may set off chains of transmission among new hosts- humans are incredibly effective vectors for this virus. This is partly due to biology - humans harbour the receptor necessary for the virus to efficiently enter cells in our respiratory tracts, where it replicates and then spreads to other hosts. It is also partly due to behaviour; global travel initially catalyzed the rapid and large-scale spread of the disease now known as COVID-19.

Viral adaptations don't necessarily stop after viruses jump a species barrier. Like many other RNA (ribonucleic acid) viruses, SARS-CoV-2 has continued to change as it circulates among humans. Up until recently, this has been at the relatively steady pace of approximately 1-2 changes per month. This is more frequent than the rates of change generally observed for DNA viruses such as hepatitis B virus, but slower than what is observed for influenza A viruses. These mutations may result in different versions of SARS-CoV-2, known as variants.

To determine which variant infected a given patient, a panoramic view of the viral genome is obtained by sequencing the entire genome, from start to finish. This is called whole genome sequencing, or WGS. WGS has been done for hundreds of thousands of SARS-CoV-2 viruses; this is possible because of high throughput technology, and because of the small size of the viral genome. SARS-CoV-2 is just a fraction of the size of a human genome. For WGS, RNA is first extracted from a patient's sample; secondly, the viral

genome is amplified and tagged. A third step consists of determining the order, or sequence of nucleic acids, the genome's building blocks. These may be generated as shorter sequences that have to be coherently assembled, like words on a page to tell a story. This important step is called assembly and is followed by further analyses to interpret and translate the results. A key component of this process is ensuring the sequence is complete and accurate, not unlike checking for grammar and spelling, ensuring we have the full story (genome) and interpretation.

The vast majority of mutations arising in SARS-CoV-2 genomes are inconsequential as far as virus function is concerned, but they can be used by genomic epidemiologists to track viral transmission at regional and global scales. However, certain viruses contain genomic changes which may pose a threat to public health. These viruses are called variants of concern, or VOCs. These are versions of the virus with changes which may also serve, from the perspective of the virus, as performance enhancers of sorts. VOCs are commonly referred to by their Phylogenetic Assignment of Named Global Outbreak LINEages or PANGOLIN name, which takes into account how variants cluster genetically and epidemiologically, and include signature mutations. One of these VOCs is known by its lineage name B.1.1.7 and was initially detected in September 2020 in the United Kingdom, and subsequently associated with increased viral activity and transmission. This VOC has a number of mutations across its genome, resulting changes in the amino acid (protein building blocks) sequence in key areas such as the Spike or S protein, such as the N501Y change in the receptor binding domain, for example. A similar variant called B.1.351, first identified in South Africa in October 2020; it also harbours the N501Y mutation, in addition to the E484K mutation which has been implicated in immune escape. A third VOC, known as P.1, was identified shortly thereafter in Brazil; it also has the N501Y and E484K mutations.

A number of challenges arise from this unwelcome turn of events. The first challenge is to rapidly detect the presence of VOCs and determine the extent of their spread. An approach to VOC detection using the same type of technology used for COVID-19 testing has accelerated the identification of possible VOCs. These screening tests pick up one or two signature mutations that have been identified in VOCs, but WGS of the virus is still required for confirmation and for detection of new variants of interest. WGS previously took several weeks from sample collection to genome sequence generation due to limited capacity for complex sample processing and data analysis. Fortunately, in recent months, access to WGS has improved and timelines have shortened. This means that data that was once perceived as reference data are now potentially immediately actionable from a public safety perspective. This significant shift in the viral landscape has translated into rapid knowledge mobilization in viral genomics for public health.

The full benefits of timely SARS-CoV-2 WGS are only realized if we pursue the following: 1) sustained support for genomics, translational virology and public health to illuminate VOC blind spots and intervene; 2) enablement of rapid and reliable data sharing with public health authorities and international repositories, and 3) establish further expertise, agility and capacity to nimbly determine the significance of new mutations detected in the viral genome. Variants under investigation are VOCs if transmission or disease severity are enhanced, or if vaccine efficacy is diminished. Thus, it is critical to rapidly determine whether emerging variants bear any of these characteristics, using both experimental and epidemiological methods. Collectively, these keys to success require a coordinated and collaborative effort among virologists, immunologists, epidemiologists and public health authorities.

The days and weeks in front of us are a moving target as we anticipate an uneasy future. Where once the optimistic objective was the eradication of SARS-CoV-2, it now seems more likely that our future will be characterized by waxing and waning levels of viral activity. We will be increasingly reliant upon genomic surveillance and the potential need for vaccine updates as dictated by ongoing viral evolution. It is hard to overstate the stakes of the moment we find ourselves in, underscoring the risks of complacency. Sustained support for Canadian science and public health with a view towards resilience and preparedness will blunt the impact of this pandemic and will serve future generations well when they face the novel pathogens that are certain to emerge over the course of the Anthropocene.

This article initially appeared in the Globe and Mail on March 16, 2021.