Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Cell Systems

CellPress

Voices

How Does Large-Scale Genomic Analysis Shape Our Understanding of COVID Variants in Real Time?

Lucy van Dorp
University College London

Disentangling Dynamics

Large-scale generation and sharing of genomic data have allowed tracking of the evolution of many thousands of closely related SARS-CoV-2 lineages over the course of the pandemic. Such "genomic surveillance" has highlighted the complex dynamics underlying viral epidemics; lineages expanding, contracting, emerging, or being lost—often shaped not by the biology of the virus but by our interactions with it.

Disentangling those lineages that rise in frequency because of an intrinsic advantage is challenging. Some early warning signs may include rapid changes in frequency relative to other circulating lineages and the presence of unusual constellations of mutations—including recurrent changes or those clustered within functionally relevant regions. Such observations have recently led to the flagging of "variants of concern" and, most notably, those first detected in the UK, South Africa, and Brazil. Biological changes in prevalence may also arise via multiple (and not mutually exclusive) mechanisms including enhanced transmissibility, changes in etiology, or the ability to bypass immunity or cross-protection.

Global surveillance sequencing is vital to pick up such cases. Some of the best support for intrinsically "fitter" viruses is when similar trends are recapitulated in different settings. With studies of endemic coronaviruses suggesting some propensity for antigenic evolution, and with the new selective pressure of mass vaccination, in all likelihood there will be more lineages to flag. Assessment of the factors underlying the success of such lineages will be vital to the pandemic response moving forward.

Muki S. Shey
CIDRI-Africa; University of Cape Town

COVID-19: 501Y.V2 versus B.1.1.7

For the first 10 months of the COVID-19 pandemic, SARS-CoV-2 seemed very stable with just very few mutations that did not seem to bother scientists much. From early December to now, the 501Y.V2, B.1.1.7, and P.1 variants have been discovered in South Africa, Britain, and Brazil, respectively, which contained multiple independent and shared mutations that are more infectious than the previous variants in circulation. These new variants are now responsible for over 50% of new cases, possibly due to a stronger binding to host cells.

A mounting host response and natural selection by the virus allows mutations and new variants to develop. Genome sequencing has enabled identification of several mutations in spike protein, the virus' "protein-in-chief," and others. The discovery of 501Y.V2 tipped-off that of B.1.1.7 thanks to surveillance and the availability of genome sequencing technology in both South Africa and Britain as well as collaboration between clinicians and scientists. Further sequencing work revealed the variants may have been circulating for several months before the discovery.

As Britain currently leads the world in sequencing viruses from positive cases, more countries need to join in and increase their active genomic surveillance. Collaboration between well- and under-resourced countries with regard to genome sequencing will be key to success. The key goal now is to interrupt the transmission of the virus before further mutations and variants develop. The silver lining, however, is that current approved vaccines will still work against these variants, even though there are reports of a possible drop in efficacy.

Elodie Ghedin
LPD/NIAID; National Institutes of Health

En Garde!

The estimated mutation and evolutionary rates for SARS-CoV-2 dictate that the number of variants likely to emerge should be limited within a short period of time. However, on a regular basis, we are made aware of new variants—carrying non-synonymous mutations at functionally relevant sites—that are sweeping regions of the globe. So far, none have been a triple threat—combined faster spread, more severe disease, and immune escape. But this means we need to be on guard: detect emerging variants earlier and more rapidly test their functional potential to better inform public health responses.

The difficulty with earlier detection of new variants, before they come to dominate a geographic region, is that we do not have equal genomic surveillance across the world. It is thus difficult to assign true origin or to identify factors that promote selection. Increased and wide-spread systematic genomic surveillance would, of course, be the first step to accelerate detection. The next is sifting through deep sequence data available to capture minority variants in protein regions that are of particular interest, such as the receptor-binding domain (RBD) of the spike protein.

Recently, there has been epidemiological evidence for increased spread of two independent variant lineages that each carry, among others, a characteristic mutation in residue 501 of the spike protein. In both cases, these were variants first identified in areas of the world with sustained surveillance efforts (the UK and South Africa). Until we can deploy wide-spread genomic sequencing efforts, in the short term we can already improve preparedness by in depth scanning of variant sequence reads to quickly inform experimental testing of new mutations.

Franziska Michor
Dana-Farber Cancer Institute; Harvard University

Mapping COVID Genome Diversity

We find ourselves in the middle of a rapidly evolving situation: as COVID cases keep increasing and the virus circles the globe, novel mutations in the virus' sequence are identified. Soon we become aware of the presence of these mutations in our state. The power of sequencing enables us to track these variants in real time, map their spread within countries and across continents, and keep tabs on their rate of increase as compared to older strains to estimate changes in transmission rates. Together, this information allows us to update mitigation strategies, plan for future vaccine composition, and enable the development of novel therapeutics.

Above and beyond globally tracking disease evolution in real time, large-scale genomic sequencing also allows us to dive deeper into the evolutionary dynamics of the virus by mapping low-frequency variants within each patient. We recently utilized a genomic sequencing dataset of Austrian patients, together with careful epidemiological identification of transmission links, to determine the time evolution of low-frequency coronavirus strains within individual patients and across transmission events. This data enabled us to estimate the transmission bottleneck size—the number of virions that establish a new infection—as 1,000 SARS-CoV2 particles on average. We furthermore identified previously unknown transmission links by mapping the variant composition of patients onto transmission pairs. Large-scale sequencing analysis thus enables us to uncover new details about the biology and epidemiology of this disease as the pandemic progresses.

Eugene V. Koonin
NCBI/NLM; National Institutes of Health

SARS-CoV-2 Adapts

The COVID-19 pandemic is an unprecedented ordeal for humanity and a major challenge to health care systems worldwide. However, with over 300,000 complete SARS-CoV-2 genomes available as of this writing in January 2021, it is also an unrivaled opportunity to understand microevolution of a virus in fine details. Arguably, the principal goal is to elucidate the evolutionary regime of the virus and the selective pressures, if such exist, that cause the replacement of dominant variants by new ones. The task is hard because constructing a robust phylogeny for such a huge number of closely related genomes is a major challenge, and distinguishing signatures of selection from biases in mutational patterns, and worse, sequencing errors is another.

All caveats notwithstanding, phylogenetic analysis of SARS-CoV-2 followed by mapping all mutations to the branches of the tree reveals many recurrent amino acid substitutions, irrespective of the exact tree topology. Although it is difficult to

confidently infer the adaptive character of any particular substitution, jointly, these recurrent mutations show a clear signal of positive selection. It appears likely that adaptation of the virus involves a network of epistatic interactions within the spike and nucleocapsid proteins, but also between these structural proteins and the non-structural polyprotein. Subsets of mutations in this network form distinct signatures of virus variants that replace each other in an evolutionary process that is at least partially adaptive.

Katie Hampson
University of Glasgow

Revealing Hidden Processes

Roughly a year ago, during an undergraduate lecture on surveillance of infectious diseases, I touched on the genomics of SARS-CoV-2. The speed at which the first genome sequences were generated and shared was unparalleled—essential for diagnostic tools and fast-tracked vaccine development. In the class, I worked through a back-of-the-envelope example of how we can use the pathogen generation time and mutation rate to infer the timing of the index case and the proportion of cases detected. Never in our history have we monitored in real-time a novel pathogen from its emergence at this level of genomic resolution. Sequencing has played an unprecedented role in our understanding of where we are now, and what the challenges are going forward.

Genomics gives us a unique window allowing us to see the scale of epidemiological processes that may otherwise be hidden from us. For example, how introductions seed new outbreaks and intensify and prolong existing ones, particularly the role of those relatively rare long-distance movements that are so difficult to observe. Genomic data have given us more than hindsight. Genomics has shone a light on the mistakes in handling the pandemic so far.

At the same time, genomics has never been more timely in informing our response. Global sharing of genomic data shows us the striking convergent emergence of new variants with functionally relevant mutations, increased transmissibility, potential to evade immunity from past infections and perhaps vaccines too. Evolutionary theory told us this would happen. With genomics, we are watching this unfold in real-time and scrambling to contain them. If we are wise, genomics will allow us to adapt and future-proof our vaccines.

This year in a lecture on that same course, I talked about zoonotic transmission—about spillover and spillback. Maybe in previous lectures this might have felt like a faraway concern, but this year, everyone cares. In conversations with colleagues, I've sometimes also felt pushback on genomics. That it is an academic luxury, an endeavor that does not inform how we implement control and prevention measures. Moreover, genomics was not even in the public vocabulary. To hear political leaders on the news discussing viral lineages feels surreal. There is no going back. Genomics is now a crucial and hopefully routine and recognized part of our toolkit. But we still need to hone our communication to talk about what we learn from genomics in everyday language, to provide interpretation and meaning from this extremely technical discipline.