



# Covid-19: How unprecedented data sharing has led to faster-than-ever outbreak research

**Advances in gene sequencing have allowed scientists to trace and monitor the COVID-19 pandemic faster than any previous outbreak. However, gaps in our knowledge of how coronaviruses work has made it difficult to understand what makes the new coronavirus special.**

23 March 2020 - By IAN LE GUILLOU

When the new coronavirus (formally known as SARS-CoV-2) was identified in China in January, scientists around the world were ready to respond. The virus's entire genetic makeup, or genome, was published [online](#) within days. By comparison, during the SARS coronavirus outbreak in 2003, this took almost three months, after the disease was [originally blamed on chlamydia](#).

Advances in the technology have brought down the cost of gene sequencing significantly and the machines are now small enough to fit in the palm of your hand. This has made it easier for a large number of samples to be sequenced around the world.

'You can see from the sequences how the virus spreads, the speed at which it's spreading and estimate the number of people that are infected. As we get more and more sequences, the more and more accurate the numbers are,' said Professor Anne-Mieke Vandamme from KU Leuven, Belgium.

Next-generation sequencing, or NGS, can generate enormous amounts of data, and the challenge becomes finding ways to analyse it properly.

In 2015, Prof. Vandamme led a project called [VIROGENESIS](#) to develop new tools to help analyse and interpret the data that comes from sequencing, particularly for laboratories that were not used to dealing with sophisticated genetic analysis.

‘When we were doing the project, there were only mainly research labs that had NGS. Now everyone has NGS,’ she said.

One of the tools developed, called [Genome Detective](#), can take the raw data from the sequencing machine, filter out results from non-viruses, piece together the genome and use that to identify the virus. It does not rely on any prior guesses or hypotheses, so it can even identify viruses that have not been seen before. This was used to confirm the first case of COVID-19 in Belgium, identifying it as a SARS-related coronavirus.

### **Online sharing**

The power of gene sequencing comes from comparing the results across different cases. Prof. Vandamme says that it has been ‘fantastic’ to see the level of collaboration internationally: ‘There is a lot more online sharing of data and sequences ... compared to the past because we have a lot more online sharing tools available.’

One of these tools is [NextStrain](#), an online resource that uses genome data to monitor the evolution of disease-causing organisms such as viruses in real time. It has tracked several outbreaks including Zika, Ebola and Dengue and has even been used to inform World Health Organization policy on seasonal flu.

Research papers typically take months to be published – an aeon in the current race to tackle the pandemic. The need to share information quickly has encouraged greater sharing of ‘preprints’, drafts of papers that have not yet been through peer review.

‘The push towards open science, open data and preprinting has really changed the way we experience the scientific discourse in this outbreak compared to previous ones,’ said Professor Richard Neher, from the University of Basel, Switzerland, who leads the NextStrain project.

NextStrain already has over 700 genomes of the new coronavirus, which it can use to trace the outbreak by detecting new mutations in the virus. The mutations do not necessarily affect how the virus behaves, but they can act as a genetic signature to link cases that are related. Like tracing your ancestry through a DNA test, a virus sequenced in Madrid, for instance, could have mutations that suggest it originated from an outbreak in Italy.

‘In the current pandemic, it gives us a lower bound on how often the virus has been introduced to a specific location,’ Prof. Neher said.

NextStrain publishes a weekly situation report that analyses these trends. The team was able to estimate that the [outbreak in Iran](#) may have been introduced by a single person, whereas at least four different introductions were responsible for the outbreak in the UK, as of 13 March.

‘(Sequencing cases) will become even more important because as we start cracking down on (the pandemic), which we hopefully will achieve, it will tell us how many transmission chains are still circulating and whether the virus is being transported from one region to another,’ said Prof. Neher.

He believes that, as the virus continues to spread, it will accumulate more genetic diversity and it will give us more information on how the virus is being transmitted.

### **Genetic blueprint**

Despite the genetic blueprint of the new coronavirus being readily available, it still does not tell us very much about how it differs from other coronaviruses. Much of what we know has come from seeing how it has spread through the population. It is now clear how different it is to previous coronavirus outbreaks, such as SARS and MERS.

'They were certainly much less easy to transmit, and also had a very different presentation in that only a few people were asymptomatic. One of the many challenges that we are facing here is that people that have only very mild symptoms have been substantial in transmitting this virus,' said Prof. Neher.

'That is much harder to control because you have to convince somebody who is basically healthy to distance themselves from others.'

Yet, it is not clear why that is the case. The traits of the virus, such as its infectiousness and severity, are driven by its proteins that are responsible for invading our cells and replicating the viral genome.

'Sequencing a genome these days is pretty fast, but for proteins it's different,' said Dr Charlotte Uetrecht, from the Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Germany. She studies coronavirus proteins through a project called [SPOCKS MS](#).

'My lab is producing the proteins (of the new coronavirus) right now. So we want to see whether they behave the same (as other coronaviruses). We usually need to produce the proteins and purify them to a certain extent so we can look at them. So it's a lot more laborious than sequencing.'

Even small changes to the viral proteins can significantly influence how they interact with each other. Dr Uetrecht studies these fleeting associations, which are crucial for the virus to replicate.

'We know a bit about how that looks, but we don't really understand which of the proteins need to associate for a new genome to be produced,' she said.

Although understanding these processes could provide new targets for antiviral drugs, Dr Uetrecht says that historically there has been little interest in studying coronaviruses as they have had little relative impact until now.

The case numbers were low for SARS and MERS and interest fell after the outbreaks, she says. 'The common-cold-causing coronaviruses were not (considered) dangerous.'

'There was not much research into coronaviruses at all, until SARS. I know a few people who have been working on coronaviruses since the '90s, and they were not very well regarded – they had a hard time getting funding. It was considered a boring, irrelevant virus.

'Now, it is very interesting again.'

*The research in this article is supported by the EU. If you liked this article, please consider sharing it on social media.*

## More info

[NextStrain](#)

[Genome Detective](#)

[SPOCKS MS](#)

[VIROGENESIS](#)

[New EU research projects and initiatives to tackle the spread of coronavirus and preparedness for other outbreaks](#)

[VIDEO: Tracking the coronavirus pandemic](#)