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.

Current Biology

CellPress

Magazine

Feature

Pandemic genomics

The ongoing global COVID-19 pandemic calls for unprecedented efforts from the biosciences. In return it offers the opportunity to study an emerging pathogen and follow its spread and adaptation to the human host in real time. So far, the virus SARS-CoV-2 has only changed very subtly, which is good news for vaccination hopes. **Michael Gross** reports.

The year 2020 has been defined by the COVID-19 pandemic, which challenged societies and political leadership around the world. It also inspired scientists to focus resources on the race to understand, treat and prevent the new disease.

In the process, enormous amounts of data have accumulated. Beyond the obvious epidemiological and medical data needed on a daily basis in the defence against the disease, there are also numerous genome sequences of the virus, enabling researchers to better understand the course of the pandemic and inform the efforts made to find drugs and develop vaccines.

Emerging threat

The first genome sequences of the new coronavirus were sequenced in Wuhan in December 2019 and served to define the new pathogen. As it was found to be related to the agent of the SARS outbreak in 2002 to 2003, it was categorised as SARS-Coronavirus-2, abbreviated to SARS-CoV-2. Like the first SARS virus, it turned out to be a zoonotic disease with a natural reservoir in bats, likely transmitted to humans via another animal species, possibly pangolin (Curr. Biol. (2020) *30*, R191–R194).

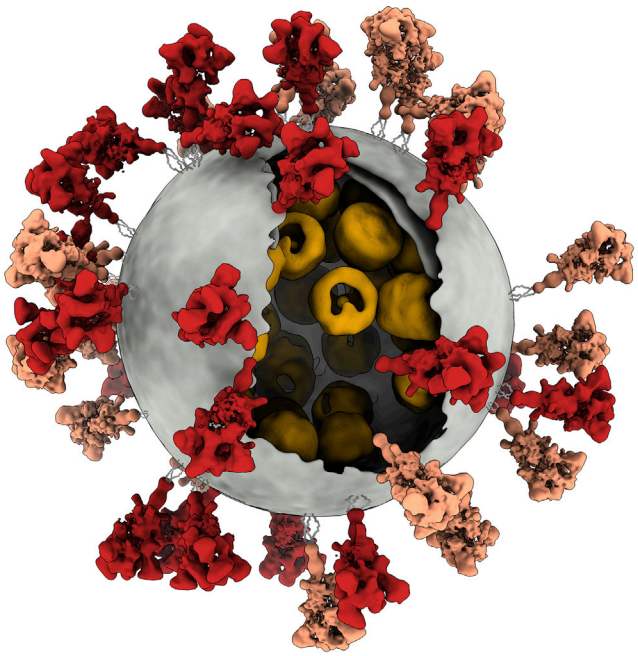
In the attempt to stop the outbreak from becoming a global pandemic, researchers traced the initial spread of the virus using a range of methods including its genome sequence. While the effort failed to rein in the disease, it did create an unprecedented trove of data on its spread and genetic variability, which researchers can now analyse in detail to better understand the disease, and the spread of pandemics in general.

The first two major analyses of the viral genomes sequenced so far appeared in September and focused on the spread of the virus in North America and in Europe. The group of Michael Worobey at the University of Arizona in Tucson, USA, used computer models based both on the genetic data and on human-related data including known transmission routes and travel behaviour (Science (2020) *370*, 564–570).

Researchers soon found that genetic data alone aren't sufficient to trace this epidemic, as SARS-CoV-2 is good at quality control in its genome replication, so it mutates less rapidly than other known RNA viruses, such as the agents of influenza and AIDS. A lasting genetic change that also translates into an amino acid replacement may happen as rarely as once a month in a given lineage. During this time the virus of one patient may have spread to hundreds of other people and travelled to other countries. Thus the virus has the natural tendency to spread faster than it changes.

Nevertheless, Worobey and colleagues were able to use the genetic information to rule out certain hypotheses about the spread of COVID-19, while reproducing others in their models and thereby making these more plausible. For instance, they could discard the suspicion that the early case detected in January at a car manufacturer in Bavaria might have been ancestral to the dramatic outbreak in Italy the following month. Instead, the analyses favour the interpretation that the Bavarian and the Italian strains of the virus made their way into Europe separately. The one that overwhelmed Italy in February went on to spread to the rest of Europe and from there to New York and Arizona.

Analyses of the spread in North America have focused on the early cases in Washington State. Both Worobey and colleagues, and an independent effort from Trevor Bedford at Fred Hutchinson Cancer Research Center in Seattle, USA,



Global threat: The coronavirus SARS-CoV-2 has spread globally and caused a pandemic on the scale of the 1918 flu pandemic. The rapid elucidation of its structural biology and genomics, however, offers opportunities to better understand its spread and to develop vaccines and treatments. (Photo: used with permission of Sai Li, Tsinghua University (Cell (2020) *183*, 730–738).)

Caged carriers: Minks bred in fur farms across Europe have turned out to be susceptible to SARS-CoV-2 and to breed new variants of the virus, which in turn might jump back to humans. (Photo: Oikeutta eläimille/Flickr (CC BY 2.0).)

(Science (2020) *370*, 571–575) have looked into this issue. Washington State identified its first case of COVID-19 on January 19, leading to early interventions that appeared to be effective at first. It was only more than a month later, from February 28, that community infections were detected.

Bedford and colleagues were able to analyse more than 10,000 samples collected between the beginning of the year and March 15 as part of the Seattle Flu Study. In these they discovered some cases that had spread before the end of February but had remained unidentified at the time. The majority of the cases in Washington State could be assigned to the same virus strain as the January 19 case, but the evidence wasn't sufficient to prove that this case was the source of the outbreak.

Taken together, these genome analyses and the ancestry trees derived from them are providing researchers with an additional tool to bridge the gaps that epidemiology faces when the virus spreads through asymptomatic carriers. They also help public health experts to better understand how the pandemic has responded to the measures taken, and how it may respond in the future.

More recently, the group of Andreas Bergthaler at the Austrian Academy of Sciences in Vienna has combined genomic analyses with the comprehensive epidemiological data available for the outbreak in Austria (Sci. Transl. Med. (2020) eabe2555). An early outbreak in the Austrian ski resort of Ischgl had notoriously acted as a superspreading event, distributing the epidemic across Europe. Epidemiologists in Iceland and Norway had identified Ischgl as a source of infections in returning travellers and alerted European authorities, but it still took weeks before the resort was closed down.

Bergthaler and colleagues analysed more than 500 genomes from the clusters in Ischgl and Vienna. They were able to reconstruct infection chains and determine important parameters, including the bottleneck population of virions in a newly infected host. They found that, typically, the population of a given patient descends from a founding population of between 10 and 1,000 virions. This is in agreement with independent work suggesting that infection typically involves the transfer of between 100 and 1,000 virus particles. This explains how imperfect measures, such as distancing and wearing non-medical masks, while not stopping all virus particles, can make important contributions to avoiding infections.

Viral variants

Most of the mutations that have so far been detected and used to track the spread of specific viral strains remain insignificant for the spread and impact of the disease. In the long term, however, the risk that a more dangerous mutation arises should be taken into consideration. The early data track how the virus spread through human societies that were naïve in immunological terms as well as, in many cases, poorly prepared in their public health policies. As societies respond to the threat with changing behaviours, vaccines and ultimately treatments, the selection pressure facing the virus may well produce new variants adapted to the changes in the host population. Like many previous zoonotic diseases, it could become attenuated but more readily transmissible and turn into just another seasonal disease like the flu. However, if it evolves to become more transmissible while remaining more lethal than the flu, it could pose further challenges.

One early mutation that has found much attention is the exchange of aspartic acid (D) for glycine (G) in position 614 of the spike protein, which decorates the outer surface of the virus and is therefore important for its interaction with host cells. The D614G variant was first discovered in April with the help of the international GISAID database, where it was flagged up as a rapidly spreading version. The Austrian study discussed above still records a mixture of both versions in the superspreading clusters analysed. Within a few months, however, the glycine version became the globally dominant one.

Whether or not this new version might make the virus more transmissible has been the subject of much scrutiny and debate. After the initial preprint publication that raised the alarm in May, Bette Korber from the Los Alamos National Laboratory, USA, and colleagues published a more comprehensive analysis in July (Cell (2020) *182*, 812–827). Based on their experiments in cell cultures, the change in the spike protein enables the virus to reproduce more rapidly in human cells than the original Wuhan strain.

The groups of Yoshihiro Kawaoka at the University of Wisconsin at Madison

and Ralph S. Baric at the University of North Carolina at Chapel Hill, both in the USA, conducted a more comprehensive study using engineered viruses in human cell cultures as well as in a hamster model. From their results published in November (Science (2020), eabe8499) the authors concluded that the "D614G substitution enhances SARS-CoV-2 infectivity, competitive fitness, and transmission in primary human cells and animal models". They also anticipate, however, that the change may make the virus more susceptible to host defences like antibodies, and this might explain why its spread hasn't impacted disease outcome.

Another rapidly spreading strain is characterised by several new mutations in addition to D614G. It appears to have emerged from Spain during the summer, when restrictions were lifted across Europe and travel resumed. Emma Hodcroft from the Biozentrum Basel, Switzerland, and colleagues have reported preliminary analyses of this variant in a preprint (medRxiv (2020) https://doi.org/10.1101/2020.10.25.20219063) released at the end of October.

As the virus continues to reproduce and spread far too easily in many countries around the world, even small probabilities of mutation add up to a serious risk. Epidemiologists worry specifically that failure to isolate patients with symptoms will favour the more virulent strains of the virus, while an efficient quarantine system might force the virus to evolve into a less conspicuous form causing milder symptoms. An additional cause for concern is the possibility of cross-infection between human patients and other mammalian hosts.

Mind your minks

Cats, dogs, lions and tigers have so far caught the virus from their human keepers but without causing too much concern. The issue that brought worries about transfer between species to the fore was the discovery of COVID-19 variants in minks (*Neovison vison*), which are bred industrially for their fur in several European countries. Denmark is the leading producer, with the mink population outnumbering the human population of the country — until the furry animals caught COVID-19.

Since June, SARS-CoV-2 has been detected in mink in more than 200 farms in Jutland, the peninsula that forms the continental part of Denmark. Genetic analyses showed that the virus changed rapidly in response to the new host, as Lucy van Dorp from University College London and colleagues reported in a preprint released on November 16 (bioRxiv (2020) https://doi.org/10.1101/2020.11.16.384743). However, these changes didn't stop the virus from jumping back to humans. Virus strains with mink-related genetic signatures were found in more than 200 human patients in Denmark alone. While most of these are likely to be comparable with the dominant strains found in human patients, there was one strain with multiple mutations that was identified as possibly more dangerous. These discoveries led to the Danish government's decree to cull the country's entire mink population.

The development worried mink farmers elsewhere and led to systematic testing. Further mink-related infections in human patients were detected in several other countries including the Netherlands, Switzerland, USA, Russia and South Africa. Most recently, COVID-19 alarms were also raised at mink farms in France and Poland.

As Francois Balloux from University College London has pointed out, even if the mink variants found so far turn out to be no more dangerous than the wild type, the presence of an additional susceptible host population held in crowded conditions and capable of breeding new viral variants is certainly an unnecessary risk that should be removed from the equation.

For this kind of risk, it is important and reassuring that the powerful methods of genome sequencing are available to monitor how the virus outbreak is evolving. Whereas earlier pandemics from the Black Death to the 1918 flu overwhelmed societies that didn't really know what hit them, genetic data give us the chance to understand what is going on and overcome the pandemic with the help of genomics.

Michael Gross is a science writer based at Oxford. He can be contacted via his web page at www.michaelgross.co.uk

Q & A

Alycia Mosley Austin

Alycia Mosley Austin, PhD, is the Interim Associate Dean of the Graduate School at the University of Rhode Island. Austin is a national leader in diversity and inclusion and oversees the Graduate School's Diversity and Inclusion Badge Program. She leads courses and workshops on professional development for trainees at all career stages. While serving as Associate Director of URI's Interdisciplinary Neuroscience Program, she managed the administrative operations of the graduate program and played a key role in developing a new undergraduate major in neuroscience. Austin holds a bachelor's degree in neuroscience from Brown University and a doctorate in neuroscience from the University of California, San Diego.

To borrow an often used term of late, do you believe that graduate schools and the scientific enterprise more generally are afflicted by 'structural racism'? Like other cultural institutions, academic science is not immune from the history of marginalization that is embedded in the foundations of our country. More importantly, academic science is not simply the product but the source of many of the ideas and practices that define structural racism. The scientific enterprise has been complicit in perpetuating the concept of race as a biological category rather than a construct created for social or political reasons. The tendency of scientists to downplay or ignore the history of how science is used to justify the dehumanization of black and brown bodies cannot be separated from the current state of the field.

When we talk about structural racism, what we are referring to is a set of actions, policies, and norms that reinforce inequities among racial groups. The fact that minoritized groups are underrepresented in science is the direct consequence of structural racism. Embedded in the graduate admissions process are assumptions about what makes a good scientist that rely on exclusionary norms. Letters of recommendation for applicants to graduate school often praise students for spending many late nights in the lab as a testament to their dedication,