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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Family tree of a deadly virus

Genetic sequencing shows the virus behind covid-19 has barely evolved, which is good news for vaccine developers, finds Graham Lawton

LIKE any other biological entity, the virus behind covid-19 has a family tree. It isn’t a very old one – SARS-CoV-2 has only been recognised since December – but it has tales to tell.

Most of what we know about this virus comes from genetic analysis. The first complete SARS-CoV-2 genome sequence was read from someone who worked at a seafood market in the Chinese city of Wuhan, who had been admitted to hospital on 26 December 2019 with symptoms of what turned out to be a new disease. Known as Wuhan-Hu-1, the sequence is a bit like the type specimen of a species – the reference against which all others are compared.

The sequence showed that the new disease was caused by a novel coronavirus closely related to a group found in bats. That was good news and bad news. Coronavirus are RNA viruses, which generally have the highest mutation rates of any known biological entity. RNA viruses are slippery customers, mutating often and hence evading drugs and immune defences.

But most coronaviruses are an exception because their RNA-copying enzyme has a proofreading function. The assumption was that the new coronavirus would follow that rule and rarely mutate.

Tens of thousands of genomes from all over the world have since been sequenced. These are constantly slotted into an ever-growing family tree by the pathogen-tracking project Nextstrain at Fred Hutchinson Cancer Research Center in Seattle.

At first all was quiet. Sequences seen in China in December and early January were identical to the reference. So were the first from outside China – three in Thailand and one in Nepal – that were analysed in January.

Then mutants started to appear. The first was collected in the Chinese province of Yunnan on 17 January. An identical one popped up in the US two days later. These were only two mutations away from the reference case, but in virology they are a new lineage.

Genetics

The SARS-CoV-2 virus has evolved minimally since December

30,000
The number of letters in the SARS-CoV-2 genome

Lineages aren’t necessarily biologically different, and that proved to be the case with this new lineage. It had no reported differences in its ability to cause an infection or its virulence. The new lineage began circulating in Asia and was soon common enough to be classed as a “clade”, which is a lineage that accounts for at least 20 per cent of cases within its branch of the family tree. As more samples were sequenced, it became clear that this new clade had actually appeared in late 2019. It was given the name 19B, to distinguish it from the original 19A clade.

For the first few weeks of 2020, these two clades – 19A and 19B – stood alone. They circulated in Asia and cropped up in North America, Europe and Australia.

In late January, a new lineage appeared in Australia and Europe. It was four mutations away from the reference genome, though again it didn’t appear to be biologically different. This lineage soon attained clade status, 20A, and dominated the European outbreaks of early 2020. It has since diverged into two other clades: 20B, which appeared in Europe, and 20C, a predominantly North American clade.

That, for now, is where we stand. All five clades constantly spawn new lineages, but none is yet common enough to become a clade. All are found worldwide, though 19A remains largely confined to Asia.

This is a lower rate of divergence than might have been expected. According to an analysis by a team at the Walter Reed Army Institute of Research in Silver Spring, Maryland, the virus has evolved to a minimal degree since December.

“The coronavirus genome is exceptionally stable,” says Samuel Díaz-Muñoz, an evolutionary virologist at the University of California, Davis. “Since the beginning of the pandemic, we’ve seen like six mutations in a 30,000-base genome. It’s one strain with minor variations.”

This is probably down to two key factors: a relatively slow mutation rate and the fact that most mutations are a hindrance to the virus and so are weeded out.

As yet there is no conclusive evidence that any of the clades has progressed to become what virologists call a “strain” – that is, a biologically different entity, perhaps a more virulent one.

One mutation, D614G, attracted attention in May after an international team warned that it appeared to be spreading rapidly, perhaps because it was more transmissible though possibly less deadly. However, that remains an unproven claim. According to the Walter Reed team, all the currently circulating viruses are sufficiently similar that a successful vaccine will immunise against them all.

But as the virus spreads, there are signs of different strains emerging. Three recently reported cases of reinfection in Hong Kong, Nevada and Belgium were caused by mutants that are all sufficiently different to evade the patients’ immune memory (see page 14). Time will tell if these constitute new strains of SARS-CoV-2.