

AD

## Covid-19: The NZ strains and our second wave of coronavirus

Marc Daalder of Newsroom · 18:50, Oct 05 2020

Daalder, 2020b

COVID-timeline  
11 August 2020

### What do studies on new coronavirus mutations tell us?

A series of studies of the genomes of thousands of samples of the new coronavirus SARS-CoV-2 show that it is mutating and evolving as it adapts to its human hosts.

2m 9s



REUTERS

This story was originally published on **newsroom.** and is republished with permission.

### ANALYSIS: What can we learn from the 29,782 nucleotides that make up the genome of the virus that caused New Zealand's second wave? Marc Daalder reports.

On July 31, a worker at the Mount Wellington branch of the Americold cold store company began to feel ill. It may have started with a dry cough, or a fever, or a loss of his sense of taste or smell.

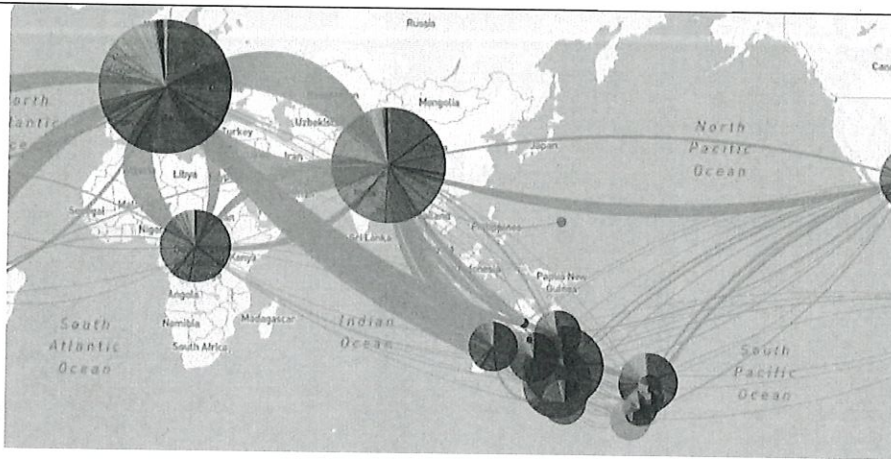
We now know that was the first sign of the reemergence of SARS-CoV-2 - the virus that causes the Covid-19 disease - in New Zealand, after months with no community cases and no local transmission. We still don't know how the virus made its way back into New Zealand, but its earliest point of entry would have been at least a day or two before the first Americold worker developed symptoms.

It wasn't until August 11, however, that Covid-19 tests of a different Americold worker and his family members came back positive. That evening, the Prime Minister announced that New Zealand's long streak of virus-free life had come to an end. Auckland was plunged into level 3 lockdown and, within a matter of weeks, the country engaged in its largest ever testing and contact tracing exercise.

More tests were conducted over the next month than in the previous three months combined. By September, masks were mandatory on public transport at level 2 and above and businesses across the country were required to display contact tracing QR codes at all alert levels.

### most popular

- 1 All Blacks vs Australia recap: Bledisloe Cup test is drawn 16-16 after a remarkable finish
- 2 Auckland apartment death: Woman fell from balcony, police say not suspicious
- 3 'The grief comes in waves': US Covid sceptic regrets throwing party that led to deaths
- 4 Election campaign: Jacinda Ardern fires up the base, Judith Collins campaigns against a ruled-out tax
- 5 Paremata stabbing victim in stable condition with serious injuries
- 6 All Blacks coach defends Rieko Ioane after star's try-line blunder proves costly
- 7 All Blacks vs Australia: Bledisloe combatants slug out a thrilling draw in Wellington wet
- 8 Newlywed who drowned at West Auckland beach 'kindhearted, strong', says grieving husband
- 9 Quiz: Afternoon trivia challenge October 11 2020
- 10 Countdown makes changes to its OneCard scheme - \$1 = 1 point



NEXTSTRAIN/SUPPLIED

This map displays the prevalence of different clades of SARS-CoV-2 on each continent.

#### READ MORE:

- \* **Covid-19: Massive genetic study shows coronavirus mutating and potentially evolving amid rapid US spread**
- \* **Coronavirus: Why New Zealand needs to ramp up genome sequencing to trace the source of its new Covid-19 outbreak**
- \* **Coronavirus: The Covid-19 mutation that has taken over the world**

Perhaps the most significant change, however, was the use of genome sequencing to help track the outbreak's progress, retroactively map its spread and link new community cases to prior ones even if an obvious path for transmission could not be identified.

Now, the genome of the virus variant that nearly brought New Zealand to its knees for a second time has been sequenced and uploaded to a public repository of SARS-CoV-2 genomes from around the world. From it, we can learn how and where the virus mutated on its way to New Zealand.

This article builds on *Newsroom's* prior investigation into the virus genomes of our first outbreak. While that is not required reading for this article, this analysis is primarily focused on the genomes of the August outbreaks and the developments in science that have occurred since the original piece was published.

#### Updated origins

SARS-CoV-2 has been circulating through animal populations in central or southwestern China for decades. One of the closest matches so far has been in a colony of horseshoe bats in China's southwestern Yunnan province - some 1900 kilometres away from the outbreak's origins in Wuhan.

MENACHEM KAHANA/GETTY IMAGES

Horseshoe bats like this have been known to carry coronaviruses.

Although that sample is 96 per cent identical to SARS-CoV-2, it was collected in 2013 and researchers don't believe it was the direct predecessor to the current human

## coronavirus

years ago.

The original SARS (76 per cent identical to SARS-CoV-2) virus has also been traced to a horseshoe bat population in Yunnan. But the 2003 SARS outbreak occurred more than 1000 kilometres away, in Guangdong province. In that scenario, scientists believe the virus made the jump to humans via Asian palm civets, small ferret- or cat-like mammals that are caught and sold live for meat in some wet markets.

Samples of coronaviruses taken from civets in wet markets in Guangzhou province, where the virus briefly reemerged in early 2004, aligned closely with the human SARS virus.

This history led researchers to initially suspect that SARS-CoV-2 might have come via a non-bat intermediary as well. The discovery of similar coronaviruses in Malayan pangolins confiscated in Wuhan, combined with the close association of Wuhan's outbreak with its wet market, fostered an early belief that the scaly mammals were the natural reservoir from which the virus emerged.

SIENA YATES/STUFF

Civets like this are thought to be responsible for spreading SARS to humans.

However, scientists have now ruled out those specific pangolins, saying that the coronaviruses they carried - while comparable to SARS-CoV-2 - were too dissimilar to be a direct ancestor.

Joep de Gijp, the head of bioinformatics and genomics at the Institute of Environmental Science and Research (ESR), our national lab testing organisation, told *Newsroom* that researchers were conducting a massive programme to sample and sequence coronaviruses from bats in China. But such a project can never be fully comprehensive.

"Just like with our current Auckland cluster, we might never find patient zero or species zero because our sampling is just not comprehensive," he said.

We also now know that the virus was likely circulating in humans for a few weeks or months prior to when it was first identified by Chinese health officials in December. That means the wet market link could have more to do with the cramped and crowded conditions of the venue for the purposes of transmission than any possible involvement in the introduction of the virus to the human population.

## Evolution

Either way, the virus has been spreading among an animal population for years, endlessly replicating itself. Each time it does so, it must accurately recreate each of the almost 30,000 nucleotides that make up the RNA - the genome - of the virus.

Each nucleotide is a sugar molecule attached to one of four different chemical bases, generally represented by a letter. In DNA, the bases are adenine (A), cytosine (C), guanine (G) and thymine (T). In RNA - like that of SARS-CoV-2 - the base uracil (U) replaces thymine.

Every once in a while, while replicating those bases, a mistake is made. This is how evolution occurs, in RNA and DNA - through accidental mutations that are then themselves replicated until a separate strain is created or the new mutation becomes dominant.

Unlike some other viruses, coronaviruses have a self-correcting mechanism that can catch most of these typos as they replicate. But eventually some slipped through.



## coronavirus

Control sequenced the RNA of the virus and uploaded it to a global flu-tracking database called GISAID on January 12.



A graphic representing the SARS-CoV-2 virus from Nextstrain. Each section contains instructions for the creation of a specific protein.

This sample, 29,903 letters long, became the official genome for the virus. It is possible - even likely - that this isn't the same virus as the one which made the original jump to the human population. Over a year, each individual letter in the virus is expected to mutate 0.0008 times. Over the length of the entire virus, that's about 23.9 mutations per year - or twice a month.

If the virus emerged in November or even October, then the official genome we use now might be a couple of mutations off. But it appears to have been prevalent enough in China at the time that it remains a reliable starting point to measure future mutations.

One of those early mutations occurred in China around January 11\*, when the virus accidentally reproduced the 241st letter, a C, as a U. By the middle of the month, the virus variant - a mutated virus is only called a strain when the mutations change its function - had tacked on two more mutations, with the 3037th letter turned from C to U and the 23,403rd letter went from A to G.



This is the D614G variant. Graphic modified from Nextstrain.

This latter change resulted in the D614G variant or strain that has been the subject of significant scientific discussion - more on that later.

## Clades

As viruses evolve, scientists work to group them into different lineages to better describe regional and genetic differences between them.

"When the virus evolves over time, it's formed different lineages on the phylogenetic tree - like a family tree - of all of these lineages," Jemma Geoghegan, an evolutionary virologist at the University of Otago, told *Newsroom*.

"Especially when you're dealing with a huge dataset, it's helpful to categorise these different lineages into a naming system. So when there's a case, for example, in New Zealand and we need to know what it's most closely related to from the global population, we don't have to look at the hundred thousand genomes that are available. We can just narrow it down to know what lineage it belongs to from that global population, which would narrow it down to maybe a couple of thousand."

These categories are called clades and three different nomenclatures for assigning SARS-CoV-2 genomes to clades have arisen.

One is that used by GISAID, the global flu-tracking database that now hosts more than 133,000 coronavirus genomes. This nomenclature is quite limited, with two top-tier clades, one of which is split into two sub-categories. One of these sub-categories is itself further split into two more categories, leaving six different possible clades overall.

Nextstrain, another open-source tool originally designed to track strains of influenza, has synthesised thousands of genomes uploaded to GISAID and estimated when and where mutations took place, as well as drawing likely connections between them. Its own nomenclature is similarly circumspect, with just five different categories for each variant to possibly fit into.

By far the most comprehensive nomenclature is called PANGOLIN (it's short for the Phylogenetic Assignment of Named Global Outbreak LINEages). This categorisation system has dozens of sub-categories and creates new ones as the virus continues to mutate.

Phylogenetic tree of the B.1.1 clade of SARS-CoV-2. The tree is rooted on the left and branches out to the right. A scale bar at the bottom left indicates the number of nucleotide changes from 0 to 5. To the right of the tree, a vertical list of accession numbers is provided, grouped by clade labels A.1, B.2, B.1, and B.1.1. The clade labels are placed in a column to the right of the accession numbers. The accession numbers are: A.1.1, A.1.3, B.10, B.14, B.13, B.18, B.3, B.9, B.5, B.6, B.4, B.7, B.2.6, B.2.2, B.2.7, B.2.4, B.2.5, B.2.1, B.1.34, B.1.6, B.1.69, B.1.71, B.1.19, B.1.23, B.1.8, B.1.22, B.1.37, B.1.13, B.1.56, B.1.66, B.1.40, B.1.12, B.1.30, B.1.31, B.1.39, B.1.34, B.1.58, B.1.41, B.1.38, B.1.3, B.1.43, B.1.29, B.1.5, B.1.33, B.1.35, B.1.72, B.1.32, B.1.67, B.1.70, and B.1.1.

The initial mutation, of the C at 241 to U, created what PANGOLIN calls the B clade and it continues to be the largest currently circulating clade. Nearly 126,500 of the 133,000 genomes currently on GISAID belong to B clade.

"The proportion of genomes sequenced around the world hugely varies from different countries," she said. The United Kingdom has uploaded 42 per cent of the genomes on GISAID, while representing under 2 per cent of the world's cases. India, by contrast, has uploaded just 2 per cent of the genomes on GISAID but makes up 18.5 per cent of the world's cases.

"This has major implications for interpreting the data. It's likely that the genomes that are omitted from the global dataset are an extraordinary amount of genetic diversity that is just not being seen."

By late January, the virus was circulating in Europe, where it experienced another mutation - the 14,408th nucleotide swapped from C to U. This launched the B.1 clade, which is the largest subcategory within the B clade - it represents 113,000 of the genomes on GISAID.

This began with the 28,881st and 28,882nd letters changing from G to A and the 28,883rd letter changing from G to C. This variant, which characterises the B.1.1 clade, would go berserk in Belgium and ultimately make its way to New Zealand as our third and fourth official cases -



## coronavirus

NEXTSTRAIN

The virus that originated in Belgium and spread throughout Europe before, ultimately, ending up in New Zealand. Graphic modified from Nextstrain.

Two more mutations came in the nine days after the B.1.1 clade was established, with the 23,731st letter changing from C to U and the 10,097th letter swapping to an A from a G. That in turn kicked off the B.1.1.1 clade, which would become extremely common in the United Kingdom - of the 6529 B.1.1.1 genomes on GISAID, 5701 of them were found in Europe. Of these, 5309 came from the UK.

Emma Hodcroft, one of the co-founders of Nextstrain and a geneticist at the University of Basel, cautioned against concluding variants are more prevalent in the United Kingdom than elsewhere.

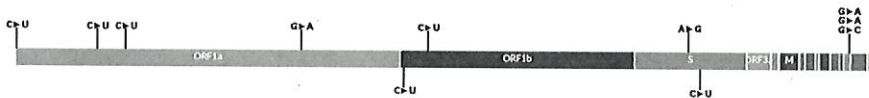
"Links to something like the UK are likely misleading," she told *Newsroom*.

"The UK simply has uploaded more sequences than anyone else - so they link to everyone."

As more genomes are uploaded from other countries, it is becoming apparent that B.1.1.1 has also spread widely in South America and Africa. About 9 per cent of the sequences from these two continents that have been uploaded to GISAID are B.1.1.1 genomes, compared to just 7.5 per cent of European genomes. In the United Kingdom, the figure is 9.5 per cent.

By March 2, the virus variant had stabilised somewhat. After the 4002nd and the 13,536th letters went from C to U, the same genome was found in Denmark on March 2 as in Ecuador on August 4 and then August 15.

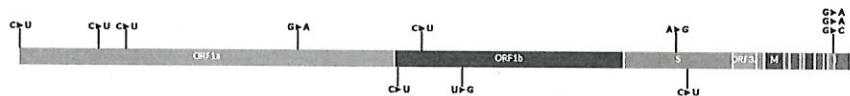
That this same variant was seen in Latin America more than five months after it first surfaced in Europe also indicates it is far more widespread than the data might suggest, ESR's de Ligt said.



NEXTSTRAIN

The B.1.1.1 clade variant seen in Denmark on March 2 and Ecuador in August. Graphic modified from Nextstrain.

Sometime in July, we now know, this same variant arrived on New Zealand's shores. Either shortly before arriving or just afterwards, another mutation occurred - the 15,867th letter swapped from U to G. This variant has so far only been found in New Zealand, but the prior mutation has yet to show up here.



NEXTSTRAIN

The B.1.1.1 clade variant unique to New Zealand. Graphic modified from Nextstrain.

It was this virus that would go on to force Auckland into level 3 lockdown for more than two weeks and plunge the entire country into a state of fear over the dreaded second wave of Covid-19.

## Reemergence

But that second wave never materialised - at least, not to the degree that it has overseas. Although the Auckland cluster now consists of 179 cases - almost double the next highest cluster - officials were able to quickly ascertain its spread and ring-fence it.

In part, that success was due to the efficacy of genome sequencing. Although genome sequencing was used to a limited extent during the first outbreak, it was quickly mobilised for the second one.

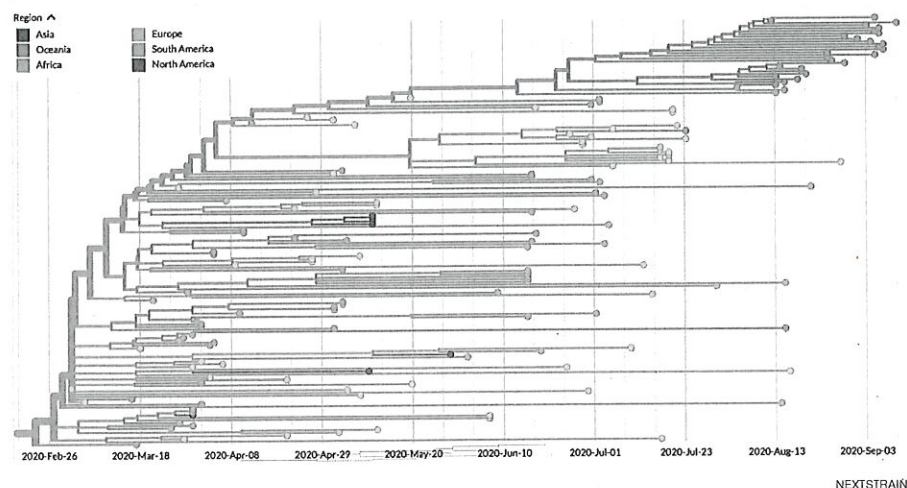
At the start of the outbreak, sequencing was used in an attempt to identify how Covid-19 returned to New Zealand. In addition to sequencing the virus genome from the new community

## coronavirus

However, some 40 per cent of recent samples were too old and degraded to successfully sequence. It remains possible that one of these samples was a match for the community variant and that this was the source of the outbreak.

"We haven't yet found an origin and that is most likely because it hasn't been sampled or sequenced," Geoghegan said.

De Ligt said he thought a border incursion - rather than human-surface-human transmission in cold storage - was the most likely culprit, but that could have occurred in managed isolation and quarantine or via sea or air ports.



The B.1.1.1 clade on Nextstrain. Each colour corresponds to a different continent - New Zealand's August outbreak is visible in the top right.

Researchers also hoped to gain an idea of how long the virus had been spreading in the country. If, for example, the virus genome showed numerous mutations that separated it from the nearest overseas variant, that could indicate it had been spreading in New Zealand for many weeks. As we have seen, however, the variant that caused the second outbreak was only one mutation away from a common overseas variant, making it likely the virus only arrived in New Zealand in the recent past.

While the early hope that genomic sequencing would reveal the origins of the outbreak has since been dispelled, the technology proved crucial in actually responding to the outbreak. On at least four occasions, new cases popped up in the community without an obvious link to the cases we had found so far.

"In the very first week or so of the recurring outbreak, the biggest question everyone had was, was this a one-off event?" de Ligt said.

"Was there one breach or were there potentially multiple breaches? What are we dealing with? Is this one lineage, one transmission chain, or is there multiple?"

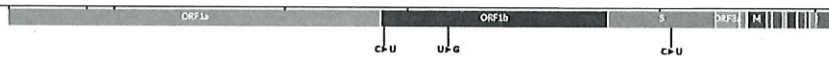
De Ligt's team sequenced the genomes of every case that popped up without an epidemiological link (that is, a viable path of transmission) to a different extant case. This reassured officials that the outbreak came from a single introduction to New Zealand. But it could also give more detailed information.

Take the case of the healthcare worker at the Jet Park quarantine facility in Auckland who tested positive for Covid-19 on September 13. The worker had no obvious link to existing community cases and could have represented either a hidden chain of transmission back to extant community cases or a breach in the quarantine facility.

ESR turned around the genomic sequencing quickly enough for the results to be reported to media the day after the case first came up.

While at least 45 of the 134 genomes from the Auckland cluster that have been uploaded to GISAID had no more mutations than the index case, the virus continued to accumulate mutations as the outbreak progressed. Generally, the later the case was found, the more mutations it had.

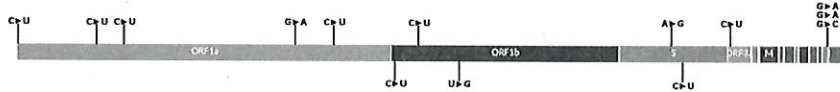
## coronavirus



NEXTSTRAIN

The original variant that kickstarted the Auckland outbreak. Graphic modified from Nextstrain.

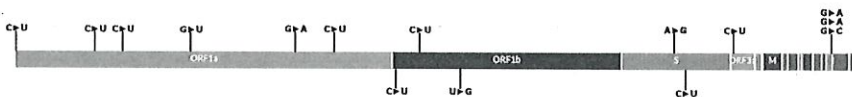
One newer mutation, in which the 11,665th letter changed from C to U, showed up in 49 of the Auckland cluster cases. Of these, 38 displayed a second mutation, a swap from C to U at position 25,685. The healthcare worker's virus genome had both of these mutations as well.



NEXTSTRAIN

Two new mutations cropped up in early August. These were present in at least 38 of Auckland's cases. Graphic modified from Nextstrain.

This was common enough. But the next mutation on the healthcare worker's genome, a change from G to U at position 6352, was shared by only three other cases - all in quarantine at the Jet Park.



NEXTSTRAIN

Just four people had the new G-to-U mutation on this genome. Graphic modified from Nextstrain.

That clue allowed de Ligt to tell health officials the exact source of the healthcare worker's infection.

"We could pinpoint, based on genomics, three people that could have infected them based on the additional mutations that were present," he said.

This situation repeated itself with the Mount Roskill church "mini-cluster", the Botany sub-cluster, the cases linked to the man who turned up to North Shore Hospital in late August with severe Covid-19 symptoms and the cases surrounding the late Dr Joe Williams.

Without genomics, these incidents might have been treated as separate clusters, raising the possibility of further escalating alert levels.

"I just want to reiterate the value of the whole genome sequencing. That's something we didn't have earlier in the year where we would have been puzzled by a number of these cases and that may have required us to give advice saying we might need to go to alert level 4 or extend longer," Director-General of Health Ashley Bloomfield said on the day Cabinet decided Auckland would move down to level 2.5.

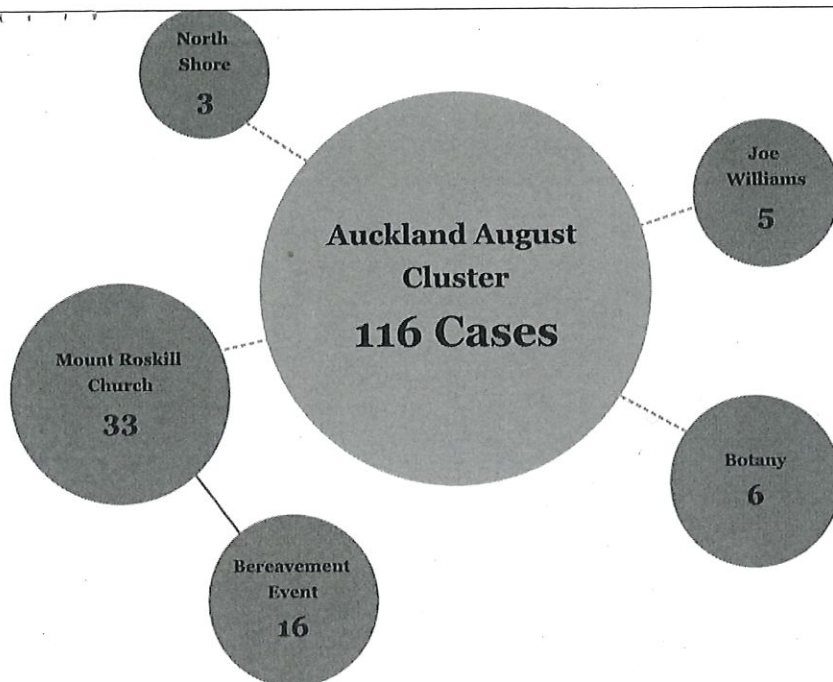
He told *Newsroom* the technology was "a game-changer".

"The whole genome sequencing has been critical in us understanding whether cases were part of the outbreak rather than novel cases, potentially from another source of infection."

Juliet Gerrard, the Prime Minister's Chief Science Advisor and a biochemist at the University of Auckland, was similarly effusive.

"There were several instances where we were reassured that seemingly unconnected clusters were actually closely linked. There were other times when the specific mutation in a sequence significantly narrowed the search for the contact tracers. Without that information, we would have been in the dark about whether new cases lacking a known link to the cluster were due to another source (e.g. a managed isolation facility) or were in fact linked to the community outbreak," she said.





"Just imagine how much harder it would have been to manage the second outbreak without this tool, with cases popping up right across Auckland."

## Strains

There is still one crucial question lingering in the minds of geneticists the world over: How are these mutations affecting the ways the virus affects us and the ways we can respond to it?

This is where the D614G variant comes in for a closer look. This particular mutation has received significant media and scientific attention in recent months, as researchers strive to figure out whether it has made the coronavirus more transmissible.

In part, some of the focus from news outlets is misguided. When countries or municipalities sequence the genomes from their outbreaks and discover the D614G variant is present, it is inevitably misreported in overseas media as a variant that is new to the world, instead of new to the region.

Although the variant likely came into existence in China in January, it has now spread to almost every country and makes up the vast majority of coronavirus cases. Of the 4687 coronavirus genomes collected in September and available on GISAID, just 16 didn't have the D614G mutation.

"One really important thing to note here is that while we learned about the mutation in spring, and while the press coverage started in summer (when publications came out), the variant actually arose early in 2020," Hodcroft told Newsroom.

"So we already know the impact of the mutation in a lot of ways - it's the variant that most people in Europe and many in the USA had in the first wave. By this I mean we do not expect this mutation to cause any differences in the autumn."

What is the impact of this mutation? It's hard to tell.

Focus first turned to it for two reasons. First, it affects the spike protein of SARS-CoV-2, which is how the virus enters cells.

Every set of three nucleotides forms an amino acid - the building blocks of proteins. Sometimes, changes to one of the three nucleotides won't change the amino acid, limiting the impact the mutation can have on the virus overall. Even when an amino acid does change, that often doesn't affect the function of the virus.

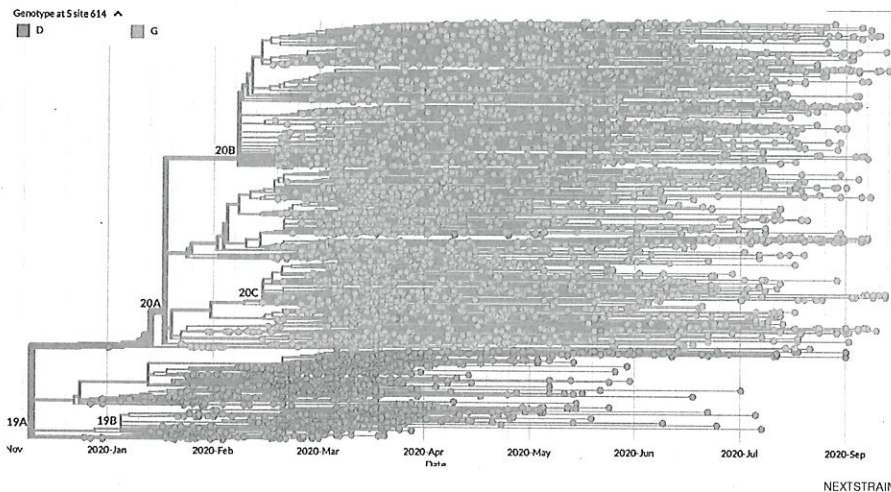
Most variants now circulating are just 10 to 12 letters different from the original genome sequenced in December of last year - and that's out of 29,903 letters. But it is possible for a mutated amino acid to have an impact.

The D614G variant arose when the 23,403rd nucleotide changed from Adenine to Guanine. That changed the 614th amino acid from aspartic acid (abbreviated with a D and made up of

## coronavirus

affecting the spike that haven't received the same level of attention. That's because D614G has spread so much more widely than almost any other mutation, to the point where it is now the dominant variant of the virus.

An April paper noting the rapid spread of the variant was criticised by other scientists for implying that the spread was a result of the mutation itself, as opposed to the environment. D614G evolved as Europe was on the cusp of mass outbreaks in Italy, Spain, France, Belgium and the United Kingdom. It is hard to tell whether it became the dominant strain in these outbreaks by coincidence or causation.



A phylogeny tree of global genomes. Yellow dots represent genomes with the D614G mutation, grey dots are those without it.

"Evidence is mounting that the mutation may lead to increased transmission, but it is hard to separate out how much of the dynamics we have seen are due to the mutation [as opposed to] other things happening at the same time," Hodcroft said.

"It's hard to separate the founder effects of this mutation with any potential selection for this mutation," Geoghegan said.

"The first cases that seeded the European outbreak in Italy had this mutation. So it's hard to know whether or not this is actually an effect of the virus with this mutation getting lucky and taking off."

Once D614G had established itself as the dominant strain in Europe, however, it was then exported to the rest of the world in vast enough numbers to quickly take over.

Scientists have tested the mutation in cell cultures from human lungs and airways and found that it can transmit as much as 10 times more easily than other variants. But it remains unclear whether that would apply to the much more complicated real life, where it has to deal with public health controls and all the idiosyncrasies of passing through a human population.

"That's in an assay, in a lab, not in living people," de Ligt said.

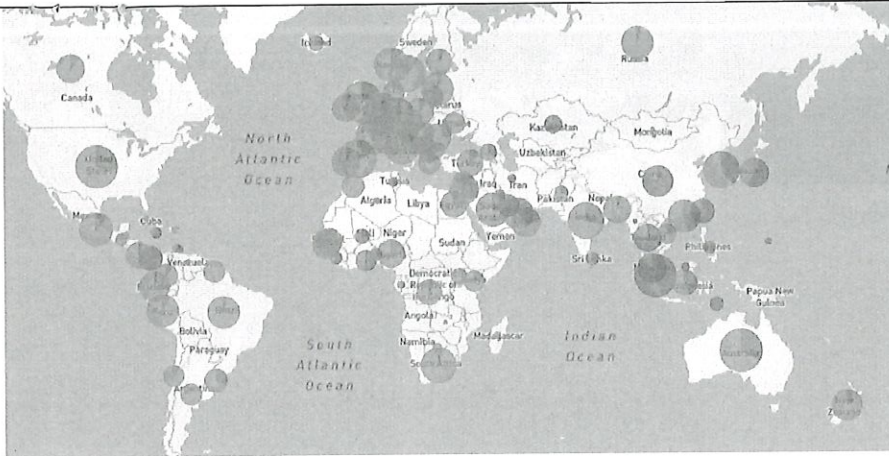
While D614G may have an impact on transmission, there remains no common variant that has affected the severity of an infection or the ability of the virus to evade the immune system. Laboratory experiments on the spike protein have found that the vast majority of potential mutations would have no effect on its ability to resist antibodies or would weaken it. Only a handful would strengthen its resistance to the human immune system.

Some of these mutations have been found in virus genomes sequenced from patients, but they are not common. Scientists believe the virus isn't facing the pressures of natural selection because it is so successfully attacking a vulnerable population that has no immunity to it.

"At the moment, there's vast populations with no immunity and the virus is doing pretty well at infecting people, so its fitness or its ability to do that is pretty good," Geoghegan said.



## coronavirus



NEXTSTRAIN

The prevalence of the D614G mutation in each country. As above, yellow represents genomes with the D614G mutation, grey represents those without.

As immunity builds up through mass infection or, more likely, vaccination, that could put more pressure on the virus to mutate to become more transmissible and better able to pierce the immune system.

"As we increase immunity, then it would have a selection pressure to work against. That's when we might see an increase in the transmission rate. But it's hard to predict whether that's going to happen."

Geoghegan said the virus will select for whatever allows it to transmit more. This could make it more severe, less severe, or have no effect on its severity from a clinical perspective - all have happened with previous viruses.

In an audit of New Zealand's genome sequencing programme, EPA chief science advisor Michael Bunce recommended the Government prioritise genomic surveillance of the virus to monitor potentially dangerous mutations.

"Aotearoa New Zealand needs to embrace genomic tools and analyses for long-term monitoring of viral evolution. This is not simply an academic exercise, rather there is a pressing need to monitor the viral lineages that are circulating (akin to seasonal influenza tracking)," he wrote.

But the slow rate of mutation - just 24 a year on average - should remain a solace both to researchers working on vaccines and the rest of us, who hope to one day take one.

**\* Nextstrain estimates are based on the composition of its dataset. As new genomes are added, some of the dates and figures in this article may no longer match what is on the Nextstrain website, but they are accurate as of time of writing.**

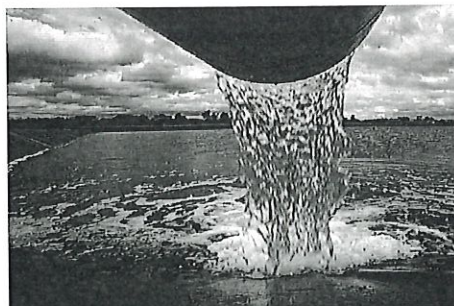
This story was originally published on **newsroom.** and is republished with permission.

Newsroom

## more from stuff



sponsored





paid content

about this content



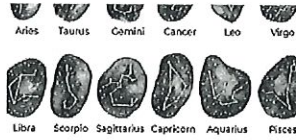
Un paso más en  
medicina digital.  
Sanitas

¡Descúbrelo!

Why isn't mindfulness  
working for me?  
www.aia.co.nz

Read more

Your Horoscope: so accurate that  
it will give you goosebumps  
Esmeralda Psychic



AD

Neighbourly | Death Notices | Coupons  
Advertising | Careers | Privacy Policy  
Cookies Policy | Terms & Conditions | Editorial Code  
Contact Us

**Breaking news?**

Send your photos, videos and tip-offs to  
newstips@stuff.co.nz, or call us on 0800 697 8833