



People in Johannesburg, South Africa, near where cases of Omicron were first identified.

# THE HUNT FOR THE ORIGINS OF OMICRON

Researchers have three key theories about how the highly transmissible variant arose. **By Smriti Mallapaty**

**L**ittle more than two months after it was first spotted in South Africa, the Omicron variant of the coronavirus SARS-CoV-2 has spread around the world faster than any previous versions. Scientists have tracked it in more than 120 countries, but remain puzzled by a key question: where did Omicron come from?

There's no transparent path of transmission linking Omicron to its predecessors. Instead, the variant has an unusual array of mutations, which it evolved entirely outside the view of researchers. Omicron is so different from earlier variants, such as Alpha and Delta, that evolutionary virologists estimate its closest-known genetic ancestor probably dates back to more than a year ago, some time after

mid-2020 (ref. 1). "It just came out of nowhere," says Darren Martin, a computational biologist at the University of Cape Town, South Africa.

The question of Omicron's origins is of more than academic importance. Working out under what conditions this highly transmissible variant arose might help scientists to understand the risk of new variants emerging, and suggest steps to minimize it, says Angela Rasmussen, a virologist at the University of Saskatchewan Vaccine and Infectious Disease Organization in Saskatoon, Canada. "It's very difficult to try to mitigate a risk that you can't even remotely wrap your head around," she says.

The World Health Organization's recently formed Scientific Advisory Group for the Origins of Novel Pathogens (SAGO) met in January to discuss Omicron's origins. The group is

expected to release a report in early February, according to Marietjie Venter, a medical virologist at the University of Pretoria in South Africa, who chairs SAGO.

Ahead of that report, scientists are investigating three theories. Although researchers have sequenced millions of SARS-CoV-2 genomes, they might simply have missed a series of mutations that eventually led to Omicron. Alternatively, the variant might have evolved mutations in one person, as part of a long-term infection. Or it could have emerged unseen in other animal hosts, such as mice or rats.

For now, whichever idea a researcher favours "often comes down to gut feeling rather than any sort of principled argument", says Richard Neher, a computational biologist at the University of Basel in Switzerland. "They are all fair game," says Jinal Bhiman, a medical scientist at the National Institute for Communicable Diseases in Johannesburg, South Africa. "Everyone has their favourite hypothesis."

## Craziest genome

Researchers agree that Omicron is a recent arrival. It was first detected in South Africa and Botswana in early November 2021 (see 'Omicron takeover'); retrospective testing has since found earlier samples from individuals in England on 1 and 3 November, and in South Africa, Nigeria and the United States on 2 November. An analysis of the mutation rate in hundreds of sequenced genomes, and of how quickly the virus had spread through populations by December, dates its emergence to not long before that – around the end of September or early October last year<sup>2</sup>. In southern Africa, Omicron probably spread from the dense urban province of Gauteng, between Johannesburg and Pretoria, to other provinces and to neighbouring Botswana.

But because Johannesburg is home to the largest airport on the African continent, the variant could have emerged anywhere in the world – merely being picked up in South Africa because of the country's sophisticated genetic surveillance, says Tulio de Oliveira, a bioinformatician at the University of KwaZulu-Natal in Durban and at Stellenbosch University's Centre for Epidemic Response and Innovation, who has led South Africa's efforts to track viral variants, including Omicron.

What stands out about Omicron is its remarkable number of mutations. Martin heard about it when he took a phone call from de Oliveira, who asked him to look at the craziest SARS-CoV-2 genome he had ever seen.

The variant has more than 50 mutations when compared with the original SARS-CoV-2 virus isolated in Wuhan, China (see [go.nature.com/32utxva](https://go.nature.com/32utxva)). Some 30 of these contribute to changes in amino acids in the spike protein<sup>3</sup>, which the coronavirus uses to attach to and fuse with cells. Previous variants of concern have had no more than ten such spike

mutations. “That is a hell of a lot of changes,” says Neher.

Researchers have seen many of these mutations before. Some were previously known to give the virus an increased ability to bind to the ACE2 receptor protein – which adorns host cells and is the docking point for SARS-CoV-2 – or to help it evade the body’s immune system. Omicron forms a stronger grip on ACE2 than do previously seen variants<sup>3</sup>. It is also better at evading the virus-blocking ‘neutralizing’ antibodies<sup>4</sup> produced by people who have been vaccinated, or who have been infected with earlier variants. Other changes in the spike protein seem to have modified how Omicron enters cells: it appears to be less adept at fusing directly with the cell’s membrane, and instead tends to gain entry after being engulfed in an endosome (a lipid-surrounded bubble)<sup>3</sup>.

But more than a dozen of Omicron’s mutations are extremely rare: some have not been seen at all before, and others have popped up but disappeared again quickly, presumably because they gave the virus a disadvantage<sup>1</sup>.

Another curious feature of Omicron is that, from a genomic viewpoint, it consists of three distinct sublineages (called BA.1, BA.2 and BA.3) that all seem to have emerged at around the same time – two of which have taken off globally (see ‘Most mutated’). That means Omicron had time to diversify before scientists noticed it. Any theory about its origins has to take this feature into account, as well as the number of mutations, notes Joel Wertheim, a molecular epidemiologist at the University of California, San Diego.

## Silent spread

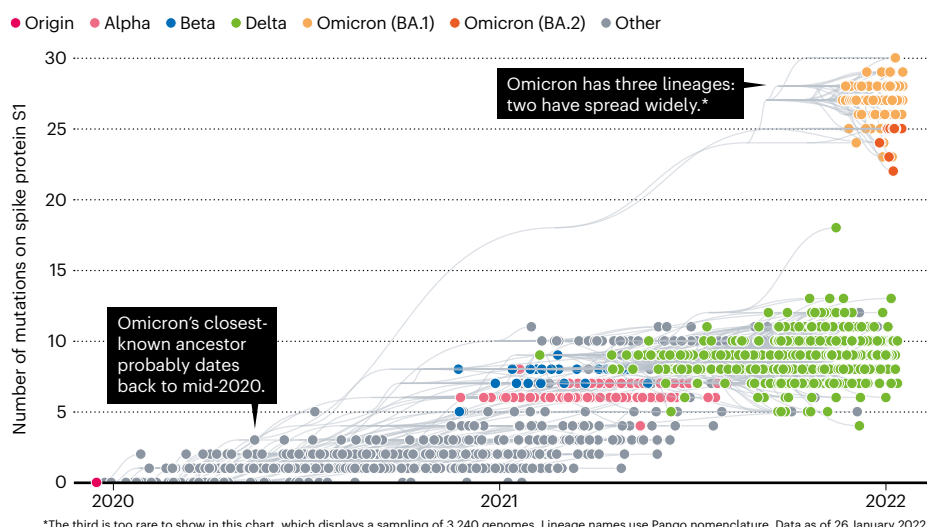
Researchers have explained the emergence of previous variants of concern through a simple process of gradual evolution. As SARS-CoV-2 replicates and transmits from person to person, random changes crop up in its RNA sequence, some of which persist. Scientists have observed that, in a given lineage, about one or two single-letter mutations a month make it into the general viral circulation – a mutation rate about half that of influenza. It is also possible for chunks of coronavirus genomes to shuffle and recombine wholesale, adds Kristian Andersen, an infectious-disease researcher at Scripps Research in La Jolla, California. And viruses can evolve faster when there is selection pressure, he says, because mutations are more likely to stick around if they give the virus an increased ability to propagate under certain environmental conditions.

Some scientists think that person-to-person spread would not be conducive to accumulating as many changes as Omicron has since mid-2020. “It does seem like a year and a half is a really short period of time for that many mutations to emerge and to apparently be selected for,” says Rasmussen.

But Bhiman argues that enough time has

## MOST MUTATED

The Omicron variant of the SARS-CoV-2 coronavirus has more mutations than any known predecessor. This chart shows mutations in the S1 subunit of the spike protein, which attaches to host cells.



elapsed. She thinks the mutation process could have occurred unseen, in a region of the world that has limited genomic sequencing and among people who don’t typically get tested, perhaps because they didn’t have symptoms. At some point in the past few months, she says, something happened to help Omicron explode, maybe because the progress of other variants – such as Delta – was gradually impeded by the immunity built up from vaccination and previous infection, whereas Omicron was able to evade this barrier.

Although researchers have submitted almost 7.5 million SARS-CoV-2 sequences to the GISAID genome database, hundreds of millions of viral genomes from people with COVID-19 worldwide have not been sequenced. South Africa, with some 28,000 genomes, has sequenced less than 1% of its known COVID-19 cases, and many nearby countries, from Tanzania to Zimbabwe and Mozambique, have submitted fewer than 1,000 sequences to GISAID.

Martin says that researchers need to sequence SARS-CoV-2 genomes from these countries to get a better sense of the likelihood of unobserved evolution. It is possible that the three sublineages of Omicron each separately arrived in South Africa from a region with limited sequencing capacity, he says.

But de Oliveira says the scenario that Omicron evolved unseen through person-to-person transmission is “extremely implausible”. Intermediate steps in Omicron’s evolution should have been picked up in viral genomes from people travelling from countries that do little sequencing to those that do a lot.

**“The virus is still expanding in the evolutionary space.”**

“This is not the nineteenth century, where you take six months to go from point to point by sailboat,” says Sergei Pond, a computational evolutionary biologist at Temple University in Philadelphia, Pennsylvania.

And Andersen adds that, because some of Omicron’s mutations haven’t been seen before, the variant might have evolved in an environment not involving person-to-person chains of transmission. Some of the changes in Omicron don’t match any seen even in the broader viral group of sarbecoviruses, which includes the virus that causes severe acute respiratory syndrome (SARS). For example, one particular site on the genomes of all known sarbecoviruses encodes a serine amino acid, but a mutation in Omicron means the variant has a lysine at that position<sup>1</sup>, which changes the biochemistry of that region, Andersen says.

However, says Jesse Bloom, a viral evolutionary geneticist at the Fred Hutchinson Cancer Research Center in Seattle, Washington, SARS-CoV-2 has not yet explored all of its possibilities in people. “The virus is still expanding in the evolutionary space.”

## Chronic infection

An alternative incubator for fast-paced evolution is a person with a chronic infection. There, the virus can multiply for weeks or months, and different types of mutation can emerge to dodge the body’s immune system. Chronic infections give the virus “the opportunity to play cat and mouse with the immune system”, says Pond, who thinks it is a plausible hypothesis for Omicron’s emergence.

Such chronic infections have been observed in people with compromised immune systems who cannot easily get rid of SARS-CoV-2. For example, a December 2020 case report described a 45-year-old man with a persistent infection<sup>5</sup>. During almost five months in its host, SARS-CoV-2 accumulated close to a dozen



## Feature

amino-acid changes in its spike protein. Some researchers suggest Alpha emerged in someone with a chronic infection, because, like Omicron, it seems to have accumulated changes at an accelerated rate (see [go.nature.com/3yjk6kmh](https://go.nature.com/3yjk6kmh)).

"The virus has to change to stick around," says Ben Murrell, an interdisciplinary virologist at the Karolinska Institute in Stockholm. The receptor-binding domain, where many of Omicron's mutations are concentrated, is an easy target for antibodies, and probably comes under pressure to change in a long-term infection.

But none of the viruses from individuals with chronic infections studied so far has had the scale of mutations observed in Omicron. Achieving that would require high rates of viral replication for a long time, which would presumably make that person very unwell, says Rasmussen. "It seems like a lot of mutations for just one person."

Further complicating the picture, Omicron's properties could stem from combinations of mutations working together. For example, two mutations found in Omicron – N501Y together with Q498R – increase a variant's ability to bind to the ACE2 protein by almost 20 times, according to cell studies<sup>6</sup>. Preliminary research by Martin and his colleagues suggests that the dozen or so rare mutations in Omicron form three separate clusters, in which they seem to work together to compensate for the negative effects of any single one<sup>1</sup>.

If this is the case, it means that the virus would have to replicate sufficiently in a person's body to explore the effects of combinations of mutations – which would take longer to achieve than if it were sampling the space of possible mutations one by one.

One possibility is that multiple individuals with chronic infections were involved, or that Omicron's ancestor came from someone with a long-term infection and then spent some time in the general population before being detected. "There are a lot of open questions," says Rasmussen.

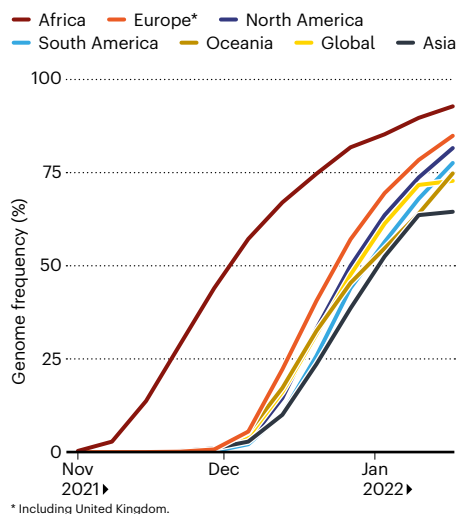
Proving this theory is close to impossible, because researchers would need to be lucky enough to find the particular person or group that could have sparked Omicron's emergence. Still, more comprehensive studies of SARS-CoV-2's evolution in chronic infections would help to map out the range of possibilities, says Neher.

### Mouse or rat

Omicron might not have emerged in a person at all. SARS-CoV-2 is a promiscuous virus: it has spread to a wild leopard, to hyenas and hippopotamuses at zoos, and into pet ferrets and hamsters. It has caused havoc in mink farms across Europe, and has infiltrated populations of white-tailed deer throughout North America. And Omicron might be able to enter a broader selection of animals. Cell-based studies have found that, unlike earlier variants,

### OMICRON TAKEOVER

Omicron has quickly spread to become the world's dominant variant of the SARS-CoV-2 coronavirus – as shown by its prevalence in viral genomes uploaded to the GISAID database from various regions.



Omicron's spike protein can bind to the ACE2 protein of turkeys, chickens and mice<sup>3,7</sup>.

One study found that the N501Y–Q498R combination of mutations allows variants to bind tightly to rat ACE2 (ref. 6). And Robert Garry, a virologist at Tulane University in New Orleans, Louisiana, notes that several other mutations in Omicron have been seen in SARS-CoV-2 viruses adapting to rodents in laboratory experiments.

The types of single-nucleotide substitution observed in Omicron's genome also seem to reflect those typically observed when coronaviruses evolve in mice, and do not match as well with the switches that are observed in coronaviruses adapting to people, according to a study of 45 mutations in Omicron<sup>8</sup>. The study noted that, in human hosts, G to U substitutions tend to occur in RNA viruses at a higher rate than C to A switches do, but that Omicron does not show this pattern.

It is possible, then, that SARS-CoV-2 could have acquired mutations that gave it access to rats – jumping from an ill person to a rat, possibly through contaminated sewage – and then spread and evolved into Omicron in that animal population. An infected rat could later have come into contact with a person, sparking the emergence of Omicron. The three sublineages of Omicron are sufficiently distinct that, according to this theory, each would represent a separate jump from animal to human.

A large population of animals with infections lasting longer than in humans could give SARS-CoV-2 room to explore a wide diversity of mutations and "build up a large ghost population of viruses that no one knows about", says Martin, who says he finds this 'reverse zoonosis' theory convincing. Changes that make the virus better at spreading in its animal host won't necessarily affect its ability to infect people, he says.

An animal reservoir could also explain why some of the mutations in Omicron have been rarely seen before in people, says Andersen.

### In the dark

But others say that even a single viral jump from an animal to a person is a rare event – let alone three. Meanwhile, the virus has had plenty of opportunities to slip between people. And although some of Omicron's mutations have been seen in rodents, that doesn't mean they can't happen or haven't occurred in people, too, and have simply been missed.

Murrell also points out that SARS-CoV-2 didn't immediately go through a period of accelerated evolution after jumping to people for the first time. When it spread to mink and deer, it did pick up changes, but not as many mutations as Omicron has accumulated, says Spyros Lytras, an evolutionary virologist at the University of Glasgow, UK. This means that the evidence isn't sufficient to suggest Omicron's predecessor would have undergone rapid selection after finding a new home in the wild.

To confirm this theory, researchers would need to find close relatives of Omicron in another animal, but they haven't been looking – "something that has been horribly neglected", says Martin. Since the pandemic began, researchers have sequenced fewer than 2,000 SARS-CoV-2 genomes isolated from other animals, mostly from mink, cats and deer.

Now that Omicron has taken off, how it evolves in people could offer more clues about its origins. It might, for instance, shed mutations that, in retrospect, are found to have helped it adapt to a different animal host, or in a person with a chronic infection. But it could also not change by much, leaving researchers in the dark.

The answer to Omicron's emergence will probably be one or a combination of the three scenarios, says Bloom. But, he adds, researchers are far from explaining the processes that brought Omicron here, let alone predicting what the next variant will look like.

And many scientists say they might never find out where Omicron came from. "Omicron really shows us the need for humility in thinking about our ability to understand the processes that are shaping the evolution of viruses like SARS-CoV-2," says Bloom.

**Smriti Mallapaty** reports for *Nature* from Sydney, Australia.

1. Martin, D. P. et al. Preprint at bioRxiv <https://doi.org/10.1101/2022.01.14.476382> (2022).
2. Viana, R. et al. *Nature* <https://doi.org/10.1038/s41586-022-04411-y> (2022).
3. Peacock, T. P. et al. Preprint at bioRxiv <https://doi.org/10.1101/2021.12.31.474653> (2022).
4. Cele, S. et al. Preprint at medRxiv <https://doi.org/10.1101/2021.12.08.21267417> (2021).
5. Choi, B. et al. *N. Engl. J. Med.* **383**, 2291–2293 (2020).
6. Bate, N. et al. Preprint at bioRxiv <https://doi.org/10.1101/2021.12.23.473975> (2021).
7. Camerini, E. et al. Preprint at bioRxiv <https://doi.org/10.1101/2021.12.12.472269> (2021).
8. Wei, C. et al. *J. Genet. Genom.* **48**, 1111–1121 (2021).