

Coronavirus's Genetics Reveal Its Global Travels

Random mutations in the SARS-CoV-2 pathogen's genome help researchers track the spread and transmission of COVID-19, the disease it causes.

Ashley Yeager

Feb 25, 2020

Several years ago, [Richard Neher](#), an evolutionary biologist at the University of Basel in Switzerland, and his colleagues wanted to monitor changes to the flu's genetic makeup to see if the data would help scientists build more-effective flu vaccines. They developed an online interface that integrated the latest viral sequencing data, analyzed it, and published the results in a publicly available [interactive web browser](#).

"Then we thought, why just flu, why not other viruses?" Neher says. The team built a similar platform to chart the transmission of MERS and Ebola and called the site [NextStrain.org](#). Now, they've adapted it to keep track of the genetic tweaks to SARS-CoV-2 as it spreads around the globe and chart the viral lineages on world maps to watch, in nearly real time, as the virus moves from major hotspots in China to smaller pockets in other countries.

Here, Neher talks with *The Scientist* about what NextStrain.org has uncovered about the transmission of SARS-CoV-2.

***The Scientist:* How do viral genomes' sequences from swabs taken from infected patients help you build a family tree of the virus?**

Richard Neher: These coronaviruses tend to change their genome, they mutate, at a fairly high rate. These mutations allow us to group viruses into more closely related viruses and less closely related viruses. All the sequences on the site are super similar because they were closely related. As time goes on, the lineages pick up independent mutations, and then they cause outbreaks in different parts of the world. You can group these sequences together by genetic makeup and reconstruct the transmission tree of the virus.

TS: Can you estimate the number of infections from the tree?

RN: Yes, if you look at the viral tree you see different

ABOVE: The viral genomes of SARS-CoV-2 reveal its path around the globe. The circles mark the size of outbreaks in different regions and the colored lines represent which strain clusters spread to those regions. The map includes data collected up to February 15. [NEXTSTRAIN.ORG](#), Leaflet, © Mapbox, © OpenStreetMap

There are 70,000 reported

sequences. And the tree will have different shapes depending on if the outbreak it's staying the same size or growing. If it's growing, you see many, many lineages coming together very deep in the tree, and that's what we have here. That implies there was rapid expansion at the base of the tree that drove all of the lineages apart. You can estimate the rate of that expansion and if you know how old the outbreak is, you can estimate the number of infections.

cases so the number of infections could be 200,000. It could be 500,000.

TS: What kind of estimates do you get using this technique?

RN: It's a little difficult to interpret the numbers from China right now. The dynamics are changing; the cases are plateauing. We expect this to be a result of these draconian containment measures or quarantine measures that they imposed on half a billion people. There are 70,000 reported cases so the number of infections could be 200,000. It could be 500,000. We don't know because people may be sick at home and stay home because the hospitals are overcrowded and that's where you could get infected. I don't think we have a good handle on how many cases there were that simply don't show up in any statistic. I would [estimate] some three-fold underreporting at least.

TS: What can the data tell you about the virus's origins?

RN: The first takeaway is that all these sequences are very, very similar, about eight mutations different than the root. That's eight mutations in a 30,000-base sequence. What this tells us is that the virus came from one source, not too long ago, somewhere between mid-November and early December.

TS: Can the mutation data tell you if the virus is becoming more virulent?

RN: We can see where the mutations change codons and where they change amino acids. Most of the mutations are probably completely inconsequential.

They just happen; it's at a rate of about one mutation per month. But we are keeping an eye on mutations that might make a difference.

There's not much doubt it will become a pandemic.

TS: How do the viral lineages help you track transmission of the disease?

RN: The mutations cluster similar transmissions together. So, families that had the virus tend to share a similar viral mutation because they had the same virus. They're a transmission cluster. So you can watch the clusters and see where they go around the world and map that spread. If you then sequence viral genomes in a new region, say, Italy, where the virus has spread, and they all were part of the same cluster, we could be reasonably confident that there was one introduction of the virus in the region. But if the viral genomes are from different clusters, that would mean there are lots of seeding events, which then make small clusters

there.

TS: Does NextStrain give you any hints about the severity of outbreaks and whether the outbreaks will become a pandemic?

RN: It doesn't really give you information on severity of outbreaks. It does tell you how different outbreaks group together and how local different outbreaks are. And so if you ask me, there's not much doubt it will become a pandemic.

Editor's note: This interview has been edited for brevity.

Ashley Yeager is an associate editor at The Scientist. Email her at ayeager@the-scientist.com. Follow her on Twitter [@AshleyJYeager](https://twitter.com/AshleyJYeager).

Keywords:

2019-nCoV, coronavirus, COVID-19, disease & medicine, flu, infection, infectious disease, influenza, News, Q&A, SARS-CoV-2, strain, transmission, vaccine