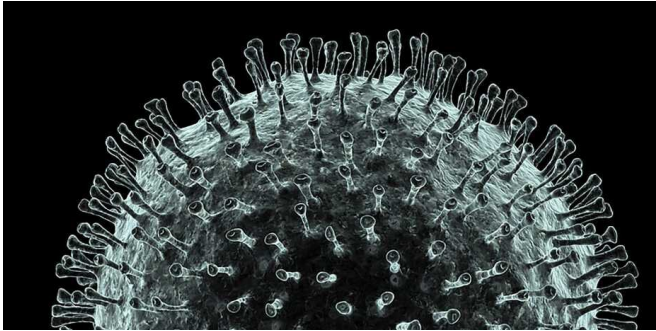


# Calculating the beginnings of the coronavirus epidemic

2 March 2020, by Fabio Bergamin



The pathogen behind the current epidemic: a coronavirus. Credit: Science Photo Library

Analyses of publicly available genome data provide clues to the beginnings of the coronavirus epidemic in China. Researchers from the Department of Biosystems Science and Engineering at ETH Zurich in Basel used a statistical model they had developed in recent years.

Since the current coronavirus epidemic started, scientists and authorities have determined the genetic fingerprint of virus samples from numerous affected countries. More than 100 of these gene sequences, which are present in coronaviruses in the form of RNA, are available in public databases. Tanja Stadler, Professor of Computational Biology at the Department of Biosystems Science and Engineering at ETH Zurich in Basel and an expert in questions of molecular epidemiology, has now studied this data. Using a [statistical model](#) her group developed to analyse the genetic genealogy of pathogens, she gained new insights into the beginnings of the epidemic in China.

Professor Stadler's analyses suggest that the epidemic in China began in the first half of November 2019, whereas most previous estimates assumed that the virus did not pass from an animal

to the first human until the second half of November. "The widespread hypothesis that the first person was infected at an animal market in November is still plausible," Stadler says. "Our data effectively rule out the scenario that the virus circulated in humans for a long time before that."

## Rapid pre-quarantine spread

Stadler also analyzed the dynamics of the epidemic before the city of Wuhan was quarantined on 23 January 2020. She used the genetic data to calculate the new coronavirus's basic reproduction number, a figure that indicates the average number of people an infected person goes on to infect. According to Stadler's estimates, it lies between two and 3.5 in the period in question. This corroborates the previous estimates based on the number of confirmed [coronavirus](#) cases, which suggested a figure between two and four. What this means is that infections occur much more quickly than with seasonal influenza (which typically has a basic reproduction number below 1.5).

"The basic reproduction number is one of the central parameters of an epidemic," Stadler says. "It provides important information on the effectiveness of measures such as quarantine. Control measures are effective only if they are able to reduce this number." That's why Stadler wants to determine what this number is during the timespan of the Wuhan quarantine. However, she says, the data for this period in Wuhan are unclear, which makes a reliable analysis impossible for now.

## Determining the hidden figure

Because viral genomes are constantly changing, Stadler could use these changes to reconstruct the evolutionary history of the virus. "Using [statistical methods](#), we can calculate how many people were infected at any point in time in the past," she explains. Her analysis showed that on 23 January, between 4,000 and 19,000 people must have been

infected. At that time there were 581 confirmed cases of the disease. This means that in the most extreme case, only one in 33 infected persons appeared in the official statistics; in the best case one in seven.

Provided by ETH Zurich

Stadler emphasizes that there are other methods than hers for determining epidemiological parameters. However, her method, which analyses the genomes, has a great advantage in that it allows reliable conclusions to be drawn even with data from relatively few patients. In particular, her method is beneficial in situations where it is no longer clear who infected whom. This is currently the case in our neighboring country Italy, and has been the case in China for some time.

### **Making real-time analysis possible**

Finally, Stadler's method would even allow the real-time analysis of an [epidemic](#), which would enable the authorities to continuously review and adjust the effectiveness of control measures. A prerequisite for this would be regular spot checks to examine the viral genome in newly infected persons. However, at present almost no sequence data is being published for new viral genomes from Wuhan.

Stadler and her colleagues have made their analysis available to other scientists on Virological, an online portal. The ETH researchers point out that their work has not been reviewed by other scientists, as is standard practice in research, as this would take too long in a situation like the current one. Stadler also stresses that the quality of her analysis can be only as good as the quality and quantity of genetic data published. In this study, her team analyzed 93 RNA sequences—most of them from China, with 38 from other countries. Stadler will continue her [analysis](#) and expand it with newly published genome data.

**More information:** Evolutionary & epidemiological analysis of 93 genomes: [virological.org/t/evolutionary... is-of-93-genomes/405](https://virological.org/t/evolutionary...-is-of-93-genomes/405)

APA citation: Calculating the beginnings of the coronavirus epidemic (2020, March 2) retrieved 1 December 2020 from <https://medicalxpress.com/news/2020-03-coronavirus-epidemic.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*