

COVID-19 may not have appeared first in China, suggests new genomic study



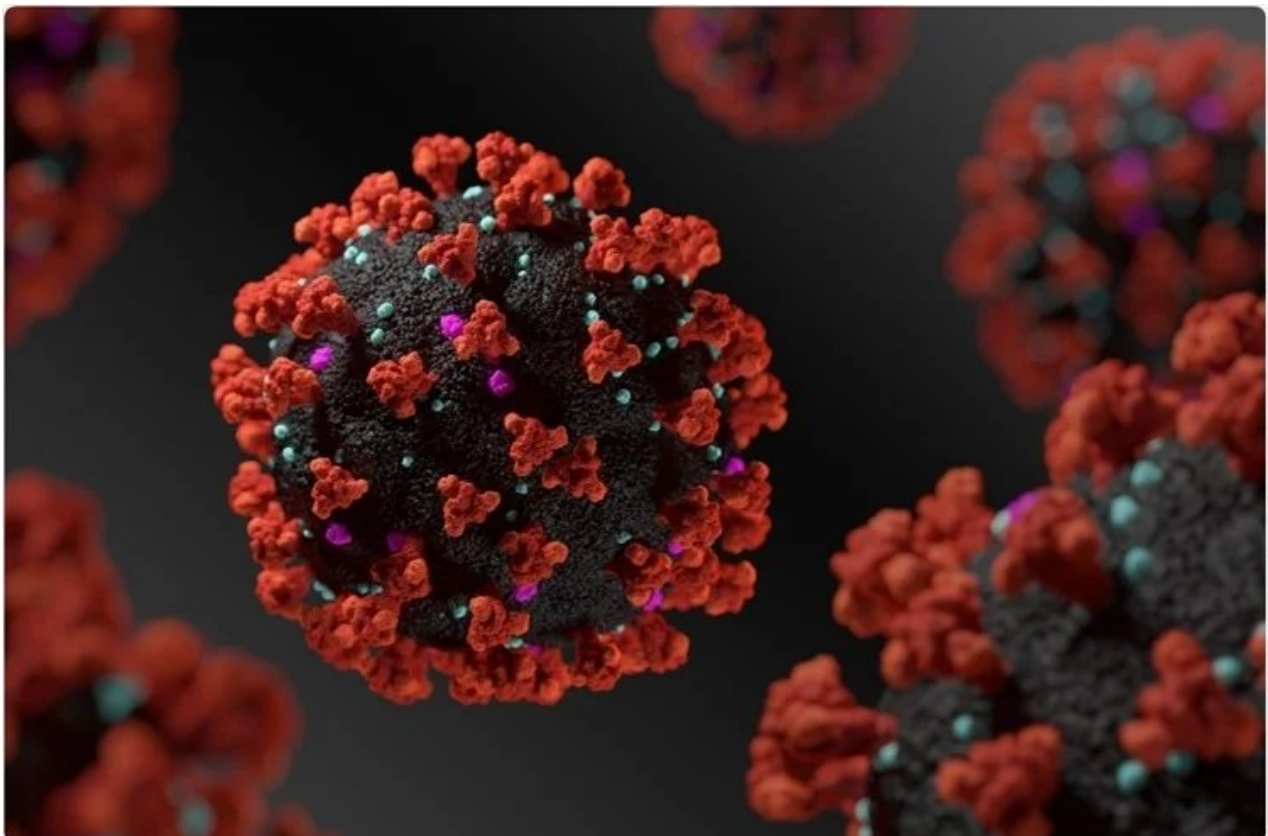
By Dr. Liji Thomas, MD

May 3 2021

The causative pathogen of coronavirus disease 2019 (COVID-19) – severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) – has led to the largest pandemic of modern times. This virus is one of seven coronaviruses known to cause human disease. While closely related to the SARS-CoV of 2002, it is far more infectious and has a long incubation period.

Thus, though it is much less deadly than the former, it has led to millions of deaths worldwide. The first case was reported from a seafood market in Wuhan, China. Since then, numerous theories have emerged regarding its origin.

A new study in *Acta Mathematica Scientia* suggests that the virus may have originated in multiple countries almost simultaneously, rather than spreading from China to the rest of the world.



Study: [Analysis of the Genomic Distance Between Bat Coronavirus RaTG13](#)

and SARS-CoV-2 Reveals Multiple Origins of COVID-19. Image Credit: SuperRune / Shutterstock

Bat coronavirus closely related

Genomic sequencing shows that the virus is most closely related to the bat coronavirus RaTG13. This seems to indicate that it sprang from a bat coronavirus lineage. The RaTG13, however, was from a 2013 sample and formed a different lineage, incapable of direct human transmission.

Many scientists have focused on collecting and sequencing samples from putative intermediate hosts, including pangolin, mink and civets, but no clear chain is observable so far.

Study aim

The earliest transmission among humans was reported from Wuhan, while other countries reported their first cases in February 2020. However, the researchers say, evidence shows that the virus was already circulating in these countries back in December 2019, including Italy, France and the USA.

In the absence of complete viral sequences from samples collected at this date in these countries, the current study hoped to examine how the currently circulating sequences of the virus may be traced back to their earliest appearance in humans.

In contrast to multiple sequence alignment (MSA), which is the conventional method of finding relationships between genomic sequences, the paper used a k-mer natural vector method to encode the complete sequence of the viral genome as vectors, based on GISAID (Global Initiative for Sharing All Influenza Data) sequences.

More accurate method

The MSA method aligns the compared sequences to obtain a matrix of similarities between them. However, such similarity fails to satisfy the triangular inequality property of mathematical distance, and so cannot show the real biological distance of different sequences.

The k-mer method encodes the vectored sequences and defines their natural distance in order to measure how close they are to each other. Whereas most studies include only a single k value to estimate distances between sequences, the current work involves all k-mers for $k \geq 1$.

They developed a new metric that satisfies the properties of positivity, non-negativity, symmetry and triangle inequality. *"The beauty of our new natural metric is that it contains information of the distributions from 1-mer to k-mer and is a mathematical metric for two genome sequences."*

Since RaTG13 was the closest in relationship to SARS-CoV-2, its distance was calculated from each of the genomes sequenced from isolates of the latter.

What were the findings?

The RaTG13 sequence was found to be closest (shortest natural distance) to those of five isolates from France, India, the Netherlands, England and the USA.

Interestingly, the viral isolates in these five cases were just as close to RaTG13 as the Wuhan isolate was. The distances with the first five were all marginally less than 31,000, which was the distance of the Wuhan isolate from RaTG13.

"These results indicate that the place where human-to-human SARS-CoV-2 transmission first happened is extremely unlikely to be Wuhan, but France, India, Netherlands, England and United States, with an accuracy rate higher than 91%."

Differences from earlier studies

Earlier studies had already suggested this possibility, since one team of scientists detected antibodies to the virus in the USA in December 2019, when no cases had been reported yet in that country. Similarly, a French study showed the presence of seropositivity (anti-SARS-CoV-2 immunoglobulin (Ig) G antibodies) in November 2019.

These studies did not include complete sequences, precluding the validation of their results by the current method. This paper advances beyond earlier uses of k-mer-based techniques by employing a one-on-one correspondence between the genome sequence and the k-mer natural vector.

Since at any value of k , the resulting k-mers will be used to calculate the newly defined metric in this study. This method conserves all available information to predict the actual biologic similarity between two sequences.

The researchers chose RaTG13 as the reference genome because it has not yet been proved that the SARS-CoV-2 reference genome (NC 045512.2) is the earliest strain. With the bat coronavirus being highly similar to the current virus, the distance from its sequence was expected to show how early the emerging strains from different countries had appeared.

What are the implications?

“Based on the results, we conclude that before the outbreak at Wuhan, China, SARS-CoV-2 most likely has already existed in other countries such as France, India, Netherland, England and United States.”

This bears out the existence of some samples that tested positive for COVID-19 before the first officially reported case in these countries.

Journal reference:

- Pei, S. et al. (2021). Analysis of the Genomic Distance Between Bat Coronavirus RaTG13 and SARS-CoV-2 Reveals Multiple Origins of COVID-19. *Acta Mathematica Scientia*. <https://doi.org/10.1007/s10473-021-0323-x>, <https://link.springer.com/article/10.1007%2Fs10473-021-0323-x>



Written by

Dr. Liji Thomas

Dr. Liji Thomas is an OB-GYN, who graduated from the Government Medical College, University of Calicut, Kerala, in 2001. Liji practiced as a full-time consultant in obstetrics/gynecology in a private hospital for a few years following her graduation. She has counseled hundreds of patients facing issues from pregnancy-related problems and infertility, and has been in charge of over 2,000 deliveries, striving always to achieve a normal delivery rather than operative.