doi:10.20944/preprints202005.0322.v1

# Understanding the origin of 'BatCoVRaTG13', a virus closest to SARS-CoV-2

Monali C. Rahalkar[1] and Rahul A. Bahulikar[2]

[1]C2, Bioenergy group, MACS Agharkar Research Institute, G.G. Agarkar Road, Pune 411004, Maharashtra, India

[2]BAIF Development Research Foundation, Central Research Station, Urulikanchan, Pune 412202

*Corresponding authors: monalirahalkar@aripune.org and bahulikar@gmail.com

## Abstract

Genomic analysis indicates that SARS-CoV-2 is most related to RaTG13, a beta corona virus derived from bats by 96% [1], a study led by Dr. Zhengli Shi. The physical description of RaTG13 is not available and at present only RaTG13 genome is available on the public database. Since RaTG13 is one of the main supports for SARS-COV-2 to have a natural origin, it is of utmost importance to understand the origin of this virus/ the exact sample from which the genome was sequenced. Here we have tried to investigate the origin of the sample from which RaTG13 would have been sequenced and its phylogenetic relationship to the previously described bat corona viruses. The genome of RaTG13 was sequenced from RNA of a bat fecal swab collected in 2013 from Yunnan, China. RNA dependent RNA polymerase (RdRp) of RaTG13 shows that it is 100% similar to that of bat corona virus BtCoV/4991 also described by Dr. Zhengli Shi and her group in 2016 [2]. BtCoV/4991 was described to be a SARS-like (SL-) corona virus from bat feces sampled in an abandoned mine from Mojiang. These facts increase the possibility that RaTG13 could be the same as BtCoV/4991, based on the RdRp sequence similarities and similarities in sample collection dates and regions. RaTG13 RNA sample could have originated from the bat feces collected from an abandoned mineshaft in Mojiang. In 2012, this mineshaft reported deaths of three miners due to pneumonia and the exact etiological reason remained unknown. Paramyxoviruses were detected in the rats of this mineshaft and in general, the conditions were filthy with bat feces covered with fungi, etc. Dr. Zhengli Shi and her coworkers studied 276 bats fecal/anal swab samples from which a single SL-corona virus was detected, and this was BtCoV/4991. Therefore, the *Rhinolophus affinis* bat fecal sample originating from the abandoned mineshaft could be the source of the RaTG13 genome.

## What is BatCoVRaTG13?

From the genome comparisons, a beta corona virus detected in a bat fecal sample (named as BatCoVRaTG13), has been proven to be the closest relative of SARS-CoV-2 [1] (96.2% overall genome sequence similarity). BatCoVRaTG13 genome sequence had been derived from a RNA sample that was obtained from the fecal sample of *Rhinolophus affinis* from Yunnan province [1]. However, BatCoVRaTG13 had not been mentioned or described before the publication describing 2019-nCoV or SARS-CoV-2 [1]. The whole genome of BatCoVRaTG13 (MN996532.1) was reported after the first sequences from SARS-CoV-2 (also known as 2019-nCoV) isolated from early patients were submitted [1]. This has been evident from the Genbank records that the genome sequence of BatCoVRaTG13. Secondly, the partial sequence RNA dependent RNA polymerase (RdRp) region of BatCoVRaTG13 suggests that it has an overall 98.7-98.9% similar to that of SARS-CoV-2 by blast analysis.

## BatCoVRaTG13 and BtCoV/4991 show the same RdRp sequence and sampling locations

However, BatCoVRaTG13 RdRp sequence when blasted showed 100% similar to the RdRp sequence of BtCoV/4991. BtCoV/4991 (KP876546) was reported by Dr. Zhengli Shi and her group in 2016 and was identified as the only SARS like (SL-) corona virus detected from the various samples analyzed from bats feces/anal swabs coming from an abandoned mineshaft from the Yunnan province, China [2]. This reference [2] has not been mentioned in the Nature publication [1] where SARS-CoV-2 was described for the first time.

In another paper published from another laboratory in Wuhan [3] approximately at the same time (January 30th, 2020) of the publication of the Nature paper[1], there is no mention of BatCoVRaTG13 and only BtCoV/4991 has been identified as the next relative in the RdRp blast [3]. This proves the fact that the BatCoVRaTG13 sequence was not online and not available to Chen et al 2020 when they submitted the paper after revision (29th January 2020) [3]. The Nature paper by Dr. Zhengli Shi [1] and co-workers was also accepted on the same day (29th January 2020) and published on 2nd February 2020.

## Do RaTG13 and RaBtCoV/4991 refer to the same RNA sample/ virus?

Now it is extremely important for the scientific community to know whether the virus referred by the name BatCoVRaTG13 by Dr. Zhengli Shi in the 2020 Nature publication [1] and described as BtCoV/4991 by Dr. Zhengli Shi in 2016 publication [2] are the same or not, or derived from the same sample or not. The geographical origins of the RNA sample from which RaTG13 was sequences is mentioned as:*Rhinolophus affinis* fecal swabs, from Pu'er city, Yunnan province. In the publication by Dr. Shi, BtCoV/4991 was also detected from *Rhinolophus affinis* fecal swab and the sample was collected from an abandoned mineshaft in Mojiang, in the Yunnan province [2]. Pu'er city and Mojiang are geographically close and these could refer to the same location. The sample collection dates for both RaTG13 (MN996532.1) and BtCoV/4991 also match, i.e. July 2013. There is no clarification of this point in the paper and neither the fact that BtCoV/4991 (or RaTG13) came from a bat fecal sample from an abandoned mineshaft nor the reference [2] is cited, which in fact the author's own work. The fact that RaTG13 and BtCoV/4991 could be referring to the same virus from the same RNA sample, has been also predicted in a preprint [4].

## Mineshaft incident in Mojiang, Yunnan, China

A mineshaft in Mojiang region, Yunnan had cases of three mineworkers suffering from pneumonia and death, which occurred in 2012 [5,6]. Regarding the same, a news was published in the reputed journal Science, which dated March 20, 2014 [5]. Scientists discovered a paramyxovirus MojV present in rats. However, the exact relationship between this paramyxovirus MojV and the deaths was not correlated [5].

## BtCoV/4991, SL-corona virus detected in abandoned mineshaft, Mojiang, Yunnan

Later on, Dr. Zhengli Shi and her group were called for the investigation of the bats and related viruses present in the mineshaft [6]. In an article published in Scientific American [6] about Dr. Zhengli Shi, referred to as the 'bat woman of China', there is an account of the same mineshaft sampling which was done in Mojiang County, Yunnan(2012-2013). Dr. Zhengli Shi also mentions that the mineshaft showed unhygienic conditions where fungus grew on the bat feces. Later on, she predicted that the fungus could be the reason of pneumonia in mine workers as per the article [6].

Dr. Shi and her colleagues did the surveillance of bat corona viruses in this abandoned mineshaft for over a year and a total of 276 bats were sampled using RT PCR approach. Six bat species were found in the mines including the horse shoe bats: *Rhinolophus sinicus* and *Rhinolophus affinis* [2]. A PCR based analysis of the RdRp analysis was done and a total of 152 partial RdRp sequences of ~400 bp were obtained. Of the 152, 150 sequences were from alpha corona viruses and only two were from beta corona viruses. Of these two, one of them was related to SL-CoV and named as (**RaBtCoV/4991**) referring to *R. affinis* sample.

## BtCoV/4991 is a unique strain representing a unique group within SARS-like coronaviruses

Phylogenetic analysis of CoV/4991 partial RdRp (KP876546.1, 370 bp) region showed that this strain of SL-CoV could be unique [2]. After phylogenetic analysis, RaBtCoV/4991 was found to represent an important and unique lineage within the SL beta corona viruses [2]. However, in the same study, the S protein of RaBtCoV/4991 was not sequenced and reported while S proteins of a few other corona viruses were sequenced [2]. This fact is rather surprising owing to the significance of the new lineage [2]. Also, even though RaBtCoV/4991 represented an important strain in the lineage, and in spite of the relatively cheap NGS facilities in current times, why the whole genome of RaBtCoV/4991 was not reported till date is still a question.

Also, there is no report mentioning about the isolation of this virus. To the best of our knowledge no further information was available about RaBtCoV/4991 apart from the RdRp region till date. However, if RaBtCoV4991 and RaTG13 are the SL- corona viruses from the same sample, why the sequencing of RaTG13 was performed after the discovery of SARS-CoV-2, and why the older sample name, BtCoV/4991 was not mentioned in Zhou et al 2020 [1] is a question.

## Unanswered questions with respect to RaTG13

The proximal origin of SARS-CoV2 has been proven or claimed to be natural [7] and one of the reasons given is that the next relative of SARS-CoV2 is BatCoVRaTG13 was from bat feces [1]. However, it would be of utmost importance for the scientific community in general if these two main questions are answered.

1. If BatCoVRaTG13[1] and BtCoV/4991 [2], refer to the same virus, why was this fact not mentioned in the first place [1]? Or are they referring to two different viruses/ samples?
2. Also, why BtCoV/4991 and the paper about the isolation of corona viruses from abandoned mineshafts have not been mentioned at all in Zhou et al 2020, when the same corresponding author had published the findings in 2016? This would have helped the readers understand the exact isolation source of RaTG13.

Although the genomic composition of SARS-CoV-2 and RaTG13 differ by 4%, it was important to report and refer to the exact sample and conditions from which RaTG13 sequence was derived. The exact sample location of RaTG13 and its similarity to BtCoV/4991 has not been mentioned in Zhou et al 2020 [1] and in no further publications till date. Understanding the origin of RaTG13, the closest genetic relative of SARS-CoV-2, would help us in understanding the origin of SARS-CoV-2. Also this would help us in developing strategies to avoid such devastating pandemics by pinpointing the conditions where zoonotic transfers can occur.

## Acknowledgments

Both the authors thank their respective institutes. No financial grants were used for the work.

## Competing Interests

The authors declare that there are no competing interests.

## Authors Contributions

MCR and RAB conceived the study. MCR wrote the manuscript and RAB performed the blast analysis and edited the manuscript. Both the authors approved the final version.

## References

1 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

2 Ge, X. Y. *et al.* Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Virol. Sin.* **31**, 31-40, doi:10.1007/s12250-016-3713-9 (2016).

3 Chen, L. *et al.* RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. *Emerg Microbes Infect* **9**, 313-319, doi:10.1080/22221751.2020.1725399 (2020).

4 Bengston, D. All journal articles evaluating the origin or epidemiology of SARS-CoV-2 that utilize the RaTG13 bat strain genomics are potentially flawed and should be retracted. *preprint* (2020).

5 Stone, R. A New Killer Virus in China? *Science* (2014).

6 Qiu, J. How China's 'Bat Woman' Hunted Down Viruses from SARS to the New Coronavirus. *Sci. Am.* **322**, 24-32 (2020).

7 Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450-452, doi:10.1038/s41591-020-0820-9 (2020).