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.

Dean Bengston (Meandering Path) https://orcid.org/0000-0002-7455-719X

Recent SARS-CoV-2 epidemiological origin studies have made their conclusion based-in-part by analyzing a bat coronavirus strain that most closely matches SARS-CoV-2 called RaTG13. However, the origins of this strain are obfuscated and therefore the genomics of the strain cannot be trusted, especially in context of determining the origin of SARS-CoV-2.

Introduction

In a recent February 2020 Nature article authored by The Wuhan Institute of Virology titled: "A pneumonia outbreak associated with a new coronavirus of probable bat origin" the authors state:

"We then found that a short region of RNA-dependent RNA polymerase (RdRp) from a bat coronavirus (BatCoV RaTG13)—which was previously detected in Rhinolophus affinis from Yunnan province—showed high sequence identity to 2019-nCoV. We carried out full-length sequencing on this RNA sample (GISAID accession number EPI_ISL_402131). Simplot analysis showed that 2019-nCoV was highly similar throughout the genome to RaTG13 (Fig. 1c), with an overall genome sequence identity of 96.2%. Using the aligned genome sequences of 2019-nCoV, RaTG13, SARS-CoV and previously reported bat SARSr-CoVs, no evidence for recombination events was detected in the genome of 2019-nCoV. Phylogenetic analysis of the full-length genome and the gene sequences of RdRp and spike (S) showed that—for all sequences—RaTG13 is the closest relative of 2019-nCoV and they form a distinct lineage from other SARSr-CoVs" (1).

That short region of RNA that they previously found was not RaTG13, but BtCoV/4991, a novel SARS-linked betacoronavirus strain (SL-CoV) that the Wuhan Institute of Virology found in a *Rhinolophus affinis* bat in an abandoned mineshaft in Yunnan in 2013 which they released as accession KP876546.1 (2), not EPI_ISL_402131. This incomplete 370 nucleotide base pair (bp) RdRp set lacked a complete genome before 2020 (3) and larger genome sets are yet publicly available. However, this is not unusual, as most of the bat coronaviruses genome accessions released in the 2013 mineshaft study are around 400bp RdRp sets, with some longer sequences GenBanks KP876505-KP876546 as their sampling method was amplification of a 440-bp fragment targeting the RdRp of all known alpha- and betacoronaviruse using RT-PCR (2). The 370bp set of BtCoV/4991 matches 100% between sequences 15322 and 15681 of RaTG13, the open-reading frame protein portion (ORF1ab) of RaTG13. No other coronavirus accessions from these studies have a 100% match to RaTG13.

Methods and Results

An analysis of the genomics between RaTG13 and BtCoV/4991 using BLAST shows a complete 100% 370bp match between sequences 15322 and 15691 of RatG13 Fig. 1 GenBank accessions MN996532.1 and KP876546.1.

Score	Expect	Identities	Gaps	Strand
684 bits(370)	0.0	370/370(100%)	0/370(0%)	Plus/Plus

```
Query  15322  GCCTCACTTGTTCTTGCTCGCAAACATACAACGTGCTGTAGCTTGTCACACCGTTTCTAT  15381
Sbjct  1      ............................................................  60

Query  15382  AGATTAGCTAATGAGTGTGCTCAAGTATTGAGTGAAATGGTCATGTGTGGCGGTTCACTA  15441
Sbjct  61     ............................................................  120

Query  15442  TATGTTAAACCAGGTGGAACCTCATCAGGAGATGCCACAACTGCTTATGCTAATAGTGTC  15501
Sbjct  121    ............................................................  180

Query  15502  TTTAACATTTGTCAAGCTGTTACGGCCAATGTTAATGCACTTTTATCTACTGATGGTAAC  15561
Sbjct  181    ............................................................  240

Query  15562  AAAATTGCCGATAAGCACGTCCGCAATTTACAACACAGACTTTATGAGTGTCTCTATAGA  15621
Sbjct  241    ............................................................  300

Query  15622  AATAGAGATGTTGACACAGACTTTGTGAATGAGTTTTACGCATATTTGCGTAAACATTTC  15681
Sbjct  301    ............................................................  360

Query  15682  TCAATGATGA  15691
Sbjct  361    ..........  370
```

Fig. 1 *BLAST analysis showing through pairwise dot identity a 100% genome match with RatG13 and BtCoV/4991, accessions* MN996532.1 *and* KP876546.

The RdRp 370bp sequence match of BtCoV/4991 and RatG13 (15322-15691) matches part of the ORF1ab protein sequences of RatG13, not the Spike (S) protein Fig. 2 (4).

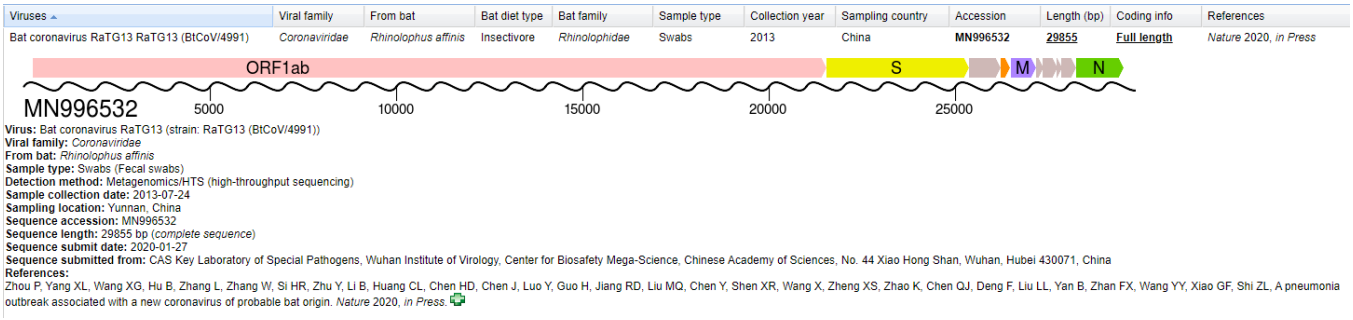

Fig. 2 *Bat coronavirus RatG13/BtCoV/4991information from the Institute of Pathogen Biology, CAMS&PUMC.*

A BLAST analysis comparison between RatG13 the 2013 mineshaft study 370bp genomic accessions KP876505-KP876546 shows that only BtCoV/4991 is identical to RatG13.

An early February 2020 study showed a 98.7% link with BtCoV/4991 and SARS-CoV-2 (3), and a BLAST analysis shows a 98.9% link with BtCoV/4991 and a Wuhan traveler with COVID-19 arriving in Korea Fig. 3 GenBank: MT039890.1.

Score	Expect	Identities	Gaps	Strand
662 bits(358)	0.0	366/370(99%)	0/370(0%)	Plus/Plus

```
Query  1      GCCTCACTTGTTCTTGCTCGCAAACATACAACGTGCTGTAGCTTGTCACACCGTTTCTAT  60
Sbjct  15340  .................................T......................  15399

Query  61     AGATTAGCTAATGAGTGTGCTCAAGTATTGAGTGAAATGGTCATGTGTGGCGGTTCACTA  120
Sbjct  15400  ............................................................  15459

Query  121    TATGTTAAACCAGGTGGAACCTCATCAGGAGATGCCACAACTGCTTATGCTAATAGTGTC  180
Sbjct  15460  ............................................................T  15519

Query  181    TTTAACATTTGTCAAGCTGTTACGGCCAATGTTAATGCACTTTTATCTACTGATGGTAAC  240
Sbjct  15520  .................C..........................................  15579

Query  241    AAAATTGCCGATAAGCACGTCCGCAATTTACAACACAGACTTTATGAGTGTCTCTATAGA  300
Sbjct  15580  ..............T.............................................  15639

Query  301    AATAGAGATGTTGACACAGACTTTGTGAATGAGTTTTACGCATATTTGCGTAAACATTTC  360
Sbjct  15640  ............................................................  15699

Query  361    TCAATGATGA  370
Sbjct  15700  ..........  15709
```

Fig. 3 *BLAST analysis showing through pairwise dot identity a 100% genome match with BtCoV/4991 and SARS-CoV-2 strain SNU01, accessions* KP876546 *and* MT039890.1.

In March of 2020, The Institute of Pathology and Biology of the Chinese Academy of Medical Sciences/Peking Union Medical College released a 2.0 accession update (4) to their Database of Bat-associated Coronaviruses, labeling the accession of RatG13, MN996532 as "strain: RatG13 (BtCoV/4991)" with the Nature RatG13 origin article as the reference with a sample collection date of 2013-07-24 from the Wuhan Institute of Virology Fig. 2 (5).

Discussion and Conclusion

In the Wuhan Institute of Virology 2013 mineshaft study where they found BtCoV/4991, the authors mention how they found a novel betacoronavirus strain linked to SARS and it is the only SARS-related strain they found in that mineshaft. In addition, they found coexistence of multiple coronaviruses in several bat species, "a phenomenon that fosters recombination and promotes the emergence of novel virus strains." (2).

In a 2018 cave study in China investigators found that SL-CoV strain ZC45 was able to infect and cause disease similar to COVID-19 in suckling rats (6), and they found that bat transmission to humans was possible without an intermediate host, such as a civet cat, via the ACE2 receptor (7). In the 2013 mineshaft study they found coexistence of multiple coronaviruses in several bat species, "a phenomenon that fosters recombination and promotes the emergence of novel virus strains." (2). No RaTG13 strain is reported in this study, the Yunnan cave studies discussed, nor in any of the Nature's origin article sources (1,2,3,6,7).

This is also unlikely to be due to a change in naming-convention, as the studies in China establishing a close link with SARS-CoV-2 and RaTG13 and the link between SARS-COV-2 and BtCoV/4991 were both published in very early February 2020 (1,3).

Therefore, these authors have obfuscated the origin of this virus, that RaTG13 was previously found in 2013 to be a SARS-linked novel betacoronavirus called BtCoV/4991 and they changed the name from BtCoV/4991 to RaTG13.

No information exists about RaTG13 before 2020. No genomics were published before then, and no information exists about RaTG13 from the 2013 Yunnan cave studies where they found and published the

genomics for BtCoV/4991 and dozens of other coronaviruses (2). The Wuhan Institute of Virology authored both of these studies; they are fully aware that RaTG13 is BtCoV/4991. If they obfuscated the origin of this virus, how can we be sure the RaTG13 genome is accurate? Their sampling technique for BtCoV/4991 amplified a 440bp portion of RdRp and we have a published 370bp set. In the Nature RaTG13 origin story, we have no information on how RaTG13 was collected and sampled. In this context we would need independent third-party access to the original specimen to extract the genetics. Science always deserves full information, not obfuscation.

Therefore, RaTG13 should otherwise be invalid for epidemiological and origin-based research studies, and studies using this genome as part of their conclusion should be retracted unless an independent third-party can verify the genomics from the original specimen.

References

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

2) Ge, X., Wang, N., Zhang, W. et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. Virol. Sin. 31, 31–40 (2016). https://doi.org/10.1007/s12250-016-3713-9

3) Chen L, Liu W, Zhang Q, *et al*. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. *Emerg Microbes Infect.* 2020;9(1):313–319. Published 2020 Feb 5. https://doi.org/10.1080/22221751.2020.1725399

4) Institute of Pathogen Biology, CAMS&PUMC. Database of Bat-associated Coronaviruses, revision document (2020). http://www.mgc.ac.cn/DBatVir/update.xml

5) Institute of Pathogen Biology, CAMS&PUMC. Database of Bat-associated Coronaviruses, RaTG13 information (2020) http://www.mgc.ac.cn/cgi-bin/DBatVir/main.cgi?func=accession&acc=MN996532

6) Hu D, Zhu C, Ai L, *et al.* Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect.* 2018;7(1):154. Published 2018 Sep 12. https://doi.org/10.1038/s41426-018-0155-5

7) Ge XY, Li JL, Yang XL, *et al*. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*. 2013;503(7477):535–538. https://doi.org/10.1038/nature12711

Supplementary Information

Web archive of The Institute of Pathology and Biology of the Chinese Academy's March 2020 2.0 Database of Bat-associated Coronaviruses revision document. http://www.mgc.ac.cn/DBatVir/update.xml

Web archive of The Institute of Pathology and Biology of the Chinese Academy's March 2020 2.0 Database of Bat-associated Coronaviruses RatG13/BtCoV4991 Information. http://www.mgc.ac.cn/cgi-bin/DBatVir/main.cgi?func=accession&acc=MN996532