Title: International expansion of a novel SARS-CoV-2 mutant

Authors: [1]Minjin Wang, MD

[1]Mengjiao Li, MD

[2,3]Ruotong Ren, PhD

[3]Lifeng Li, PhD

[4]En-Qiang Chen, PhD

[5]Weimin Li, PhD

[1]Binwu Ying, PhD

Minjin Wang, Mengjiao Li, Ruotong Ren contributed equally to this work.

Author order was determined both alphabetically and in order of increasing seniority.

Author Affiliations:

[1]Department of Laboratory Medicine, West China Hospital of Sichuan University, Chengdu, China (Minjin Wang, Mengjiao Li, Binwu Ying)

[2]Department of Hematology, The First Hospital of Lanzhou University, Lanzhou, China (Ruotong Ren)

[3]Genskey Biotechnology Co., Ltd., Beijing, China (Ruotong Ren, Lifeng Li)

[4]Center of Infectious Diseases, West China Hospital of Sichuan University, Chengdu, China (En-Qiang Chen)

[5]Department of Respiratory and Critical Care Medicine, West China Hospital of Sichuan University, Chengdu, China (Weimin Li)

Corresponding author:

Binwu Ying, PhD, Department of Laboratory Medicine, West China Hospital of Sichuan University, Chengdu, China.

Email: binwuying@126.com, Tel: (0086)85423559

Address: Number 37 Guoxue Alley in the wuhou District, Chengdu, Sichuan, China

International expansion of a novel SARS-CoV-2 mutant

RNA viruses such as coronavirus are rapidly evolving pathogens that can accumulate considerable genetic diversity in relatively short time periods. Mutation accumulated in SARS-CoV-2 genomes during its pandemic spread can cause unpredictable effects on COVID-19 and further complicate epidemic control efforts[1]. Here we report that a novel SARS-CoV-2 mutation in its ORF3a gene appears to be spreading worldwide, which deserves close attention.

We collected 95 SARS-CoV-2 samples from Sichuan Province of China for amplification-free whole genome sequencing and acquired 13 whole genome sequences, which were analyzed for sequence variation and evolution together with 199 SARS-CoV-2 genomes publicly released in the GISAID EpiFlu™ database (https://www.gisaid.org/) and 7 genomes download from NGDC database (https://bigd.big.ac.cn/ncov). This study was approved by the Biomedical Research Ethics Committee of West China Hospital of Sichuan University (reference no. 193, 2020) with a waiver of informed consent.

Based on 10 high frequency mutations (mutant allele frequency >5%), these SARS-CoV-2 genomes can be classified into 5 main groups: original stain 1 and 4 variants with different mutations groups and clusters (Figure). The most common variants (Group 1) exhibited both a missense mutation (ORF8:c.251tTa>tCa; present in 31.58% of the isolates) and a synonymous mutation (orf1ab:c.8517agC>agT; found in 30.62% of the isolates), suggesting a possible linkage between these two sites. Also, 3 subgroups were evolved in the main Group 1 by other 3 mutations. Group 2 was clustered together with 3 mutants including missense variant S: c.1841gAt>gGt, orf1ab upstream gene variant and synonymous variant orf1ab: c.2772ttC>ttT. Group 3 viral isolates were much less frequent (11.48%) and characterized by a missense mutation (orf1ab:c.10818ttG>ttT). Group 4 viral isolates contained a novel missense mutation (ORF3a:c.752gGt>gTt) first identified in a Chinese family. Notably, however, Group 4 viral isolates were most frequently found outside mainland China (23.28%; 27/116; $p<0.01$ by Fisher's exact test). Additionally, Group 2 and Group 4 showed obvious aggregation in non-Chinese countries and regions.

The family (an old female and two young family members) who carrying the Group 4 variant returned from Wuhan to their hometown in Sichuan on January 20, 2020. By January 23, the

old female exhibited symptoms of fever and cough, and her two children also developed these symptoms in the following days. Their throat swab samples were tested SARS-CoV-2 positive by reverse real-time PCR assay on January 25. The old female with chronic hypertension was in a critical condition after suffering the COVID-19 disease while the two young family members shows slight symptoms. The underlying disease might contribute to the illness progress. None of these individuals traveled outside of China between the start of the COVID-19 epidemic and their return to Sichuan, however the Group 4 variant has demonstrated global dissemination.

We performed a timeline analysis using the sample collection dates reported in the GISAID EpiFlu™ database. Except 3 patients from Sichuan, China, who traveled from Wuhan prior to their symptom onset, isolates in Group 4 with ORF3a mutant were subsequently reported in several other countries and regions, including China (Taiwan), France (Paris), and Australia (Sydney and Clayton), Singapore, South Korea, the United Kingdom and Italy. It should be noted that this mutant virus strain appears to be the most prevalent form of SARS-COV-2 in France, Italy, Brazil, and Singapore.

Virus genome data from France indicate that SARS-CoV-2 strains carrying ORF3a:c.752gGt>gTt often have a S:c.1099Gtc>Ttc mutation in their S gene, which interacts with ACE2 mediating viral entry into its host cells[3], and is regarded as a critical factor for viral transmission and virulence[4, 5]. It is not yet clear whether this mutation is common in Group 4 viral isolates from different geographical regions. Given the prevalence of Group 4 isolates in multiple countries, including France, Italy and South Korea, which is experiencing a rapidly growing epidemic, this information should be of significant importance further investigate whether this mutation enhances host cell entry.

At present, the SARS-CoV-2 epidemic in China is diminishing owing to collected control efforts, but the rapid global spread has become a major health concern. Very little is known about how rapidly the SARS-CoV-2 genome mutates and how this affects transmission or pathogenesis. Our findings indicate that comprehensive studies combining genome, epidemiological and clinical data urgent needed to clarify these issues.

Author Contributions:

Dr Ying had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Binwu Ying.

Acquisition, analysis, or interpretation of data: Minjin Wang, Mengjiao Li, Ruotong Ren, Lifeng Li, En-Qiang Chen

Drafting of the manuscript: Minjin Wang, Mengjiao Li, Ruotong Ren, Binwu Ying.

Critical revision of the manuscript for important intellectual content: Binwu Ying, Weimin Li.

Statistical analysis: Minjin Wang, Mengjiao Li, Ruotong Ren, Binwu Ying.

Obtained funding: Weimin Li, Binwu Ying.

Administrative, technical, or material support: Binwu Ying.

Supervision: Weimin Li. Binwu Ying.

Conflict of Interest Disclosures: None reported.

Funding/Support: This study was funded by Science & Technology Department of Sichuan Province (Grant number: 2020YFS0004) and West China Hospital of Sichuan University (Grant number: HX-2019-nCoV-066).

Acknowledgments: Dr. Andreas Bråve (Department of microbiology, Public Health Agency of Sweden) and Dr. Sylvie van der Werf (Department of Virology, Molecular Genetics of RNA Viruses unit, CNRS UMR-3569, University of Paris, National Reference Center for Respiratory Viruses Institute Pasteur) have provided epidemiological characteristics of infected people in the region for us to finish the work. We are very grateful for their efforts. Finally, we also thank Dr Wan Xiong for polishing the manuscript language, Shuo Guo and Yanbing Zhou for collecting literature.

References

1. www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/
2. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality[J]. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 2017,22(13). doi:org/10.2807/1560-7917.es.2017.22.13.30494
3. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses[J]. Nature microbiology. 2020. doi:org/10.1038/s41564-020-0688-y
4. Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond[J]. Trends in microbiology. 2015,23(8):468-478. doi:org/10.1016/j.tim.2015.06.003
5. Hamming I, Timens W, Bulthuis ML, Lely AT, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis[J]. The Journal of pathology. 2004,203(2):631-637. doi:org/10.1002/path.1570

Figure legend:

Figure 1: Maximum likelihood tree based on the whole genome sequences of 221 viral strains.

199 high quality genomes were collected from GISAID EpiFlu™ database, including 1 *Rhinolophus affinis* isolate, 6 *Manis javanica* isolates and 2 environmental isolates. 22 additional genomes were collected from other resource, including 7 genomes from NGDC (https://bigd.big.ac.cn/ncov), 13 genomes from West China Hospital of Sichuan University(WCH). SARS-CoV (NC_004718.3) and MERS-CoV (NC_019843.3) genomes sequence were downloaded from NCBI RefSeq database. MAFFT (version 7.543) was used for sequence alignment, and PhyML (version 3.0) was used to construct the evolutionary tree. Variation information of human SARS-CoV-2 genome was derived from NGDC. Mutations of 13 WCH genomes were analyzed using NGDC online tools (https://bigd.big.ac.cn/ncov/tool/variation-identify).