**Characterizing COVID-19 severity, epidemiology and SARS-CoV-2 genotypes in a regional business hub of China**

Yan Yan 1, 2, 3#, Bin Liu 1#, Hui Ding 1#, Xu Wang 1, Yaping Dai 1, Difei Ding 1, Hui Yu 1, Mingchao Sha 1, Chunyan Lui 1, Davgadorj Chantsalmaa 1, Yuanwang Qiu 2, 3\*, Lihua Huang 2, 3\*, Qinxue Hu 4, 5\*

1 Center of Clinical Laboratory, The Fifth People's Hospital of Wuxi (Wuxi Infectious Disease Hospital), Affiliated Hospital of Jiangnan University, Wuxi, China

2 The International Joint Research Laboratory for Infection and Immunity (China-Germany), Jiangnan University, Wuxi, China

3 Hepatology Institute of Wuxi, The Fifth People's Hospital of Wuxi (Wuxi Infectious Disease Hospital), Affiliated Hospital of Jiangnan University, Wuxi, China

4 State Key Laboratory of Virology, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, China

5 Institute for Infection and Immunity, St. George’s, University of London, London, United Kingdom

# These first authors contributed equally to this article.

\*Correspondence:

Yuanwang Qiu 306126019@qq.com

Lihua Huang [huanglihua1964@sina.com](mailto:huanglihua1964@sina.com)

Qinxue Hu [qhu@wh.iov.cn](mailto:qhu@wh.iov.cn)

Key words: COVID-19; SARS-CoV-2; disease severity, epidemiology; genotype

*To the Editor:*

We read with great interest the article of Sijia Tian et al [1](#_ENREF_1) on “Characteristics of COVID-19 infection in Beijing”, which informed the ratio of COVID-19 clinical outcome and epidemiological characteristics in Beijing. Clinical outcome-based surveillance is conducive to make rational assessment of medical cost, while epidemiology and genotypes of SARS-CoV-2 circulating locally and globally are important for prevention and vaccine development. China authority agencies have established an open and real-time updated online database of SARS-CoV-2 sequences (<http://nmdc.cn/coronavirus>) as an information resource for scientists and clinicians.[2](#_ENREF_2) However, it remains challenging to control and prevent this virus infection since information is rapidly evolving, and there are still many gaps that are not yet fully understood as to how this animal virus crossed species boundaries to infect humans.[3](#_ENREF_3), [4](#_ENREF_4) Based on epidemiological analyses and prediction, this highly contagious virus has a high probability of causing a recurrent epidemic in the 2020/2021 winter.[5](#_ENREF_5) Therefore, to better control future outbreaks, it is important to understand SARS-CoV-2 transmission and pathogenesis by characterizing the clinical and epidemiological features of COVID-19 and the SARS-CoV-2 strains circulating locally and globally.

In this retrospective study, we investigated COVID-19 disease severity, epidemiology and genotypes of SARS-CoV-2 in a regional business hub, Wuxi of China. Posterior oropharyngeal mucosal specimens of suspected patients were collected. We studied fifty-five COVID-19 patients (Ethics No. 2020-010-1) admitted to the hospital between January 25 and March 31, 2020. The clinical results showed that most cases were mild illness (45/55, 81.8%) with the rest being severe (10/55, 18.2%), and were mainly imported (Generation one cases (G1) = 35, 63.6%) and occurred mainly in January and February of 2020 after the implement of monitoring and quarantine measures. The cases in March were significantly decreased, and most source cases were from neighboring provinces (Wuhan, Henan, Anhui, and Shanghai) and foreign countries (USA, Germany, Japan, and Philippines) (Fig. 1A and 1B). High incidence of severe illness was found in G1 (6/35, 17.1%) and G2 group (4/15, 26.7%). Nearly 32.7% (18/55) cases (G2 plus G3 groups) were attributable to household transmission, and 3.6% (2/55) G2 had no clear community contact history (Fig. 1B and 1C). Among the total cases, 58.2% (32/55) were male, and men (7/32, 21.9%) also had a higher proportion of severe cases than females (3/23, 13.0%) (Fig. 1C). The main infection group was 20-44 years old (17/55, 30.9%). The proportion of severe cases with comorbidities (8/10, 80%) was ~4-fold greater than that of mild cases (10/45, 22.2%) (Fig. 1E), while 60% (6/10) severe cases occurred in over 55 years old (Fig. 1F). The average age (53.4 ± 6.6) of severely ill patients was significantly higher than all patients’ average age (41.6 ± 2.6) (*p*=0.038).

Concerning SARS-CoV-2 transmission and pathogenesis, current knowledge is limited. Coronaviruses are prone to mutations because they are single-stranded RNA viruses.[6](#_ENREF_6) It is probable that these viruses would undergo mutations over time that could substantially change their features.[6](#_ENREF_6), [7](#_ENREF_7) SARS-CoV-2 genomes of different pandemic locations inside [6](#_ENREF_6), [7](#_ENREF_7) and outside of China [8](#_ENREF_8) have become available. It is crucial to monitor protein mutations in viruses, which refers to changes in the sequence of amino acids of SARS-CoV-2 genome. According to the changes in the genomic base and amino acid sequences, the mutation time point of a virus can be traced.[9](#_ENREF_9) If the amount of data is sufficient, the development can be traced from the perspective of space and time.[2](#_ENREF_2) Such information, in combination with the clinical outcomes of infected individuals, may help design prevention and treatment strategies.

There are several standards for the classification of SARS-CoV-2 epidemic strains at home and abroad. [7](#_ENREF_7), [8](#_ENREF_8) We adopted Tang’s analysis and classification method which was developed to analyze 103 available whole-genome sequences from the early stage of the SARS-CoV-2 outbreak. [7](#_ENREF_7) According to Tang’s analysis, S-type and L-type subtypes were prevalent at the early period of the SARS-CoV-2 outbreak, and the L-type dominated the epidemic in Wuhan in the early period (~70%), but the frequency subsequently decreased since early January 2020.[7](#_ENREF_7) Tang’s study has been confirmed by Forster et al [8](#_ENREF_8), in which S-type corresponds to A-type, and L-type corresponds to B-type. This nomenclature is now being used by GISAID. We subsequently sequenced partial ORF1a and full ORF8 of seven SARS-CoV-2 Wuxi strains, which covered the highest frequency of genomic variation in ORF1a (nt 8782) and ORF8 (nt 28144), respectively. Finally, the regions of partial ORF1a (nt 8573-8938, 366 bp) and full ORF8 (nt 27881-28246, 465 bp) were identified. Given the need for a specific segment comparison with the global reference genes, we registered Wuxi strains (CHN/JS(5)/2020\_02\_02, CHN/JS(13)/2020\_02\_04, CHN/JS(16)/2020\_02\_02, CHN/JS(21)/2020\_01\_25, CHN/JS(29)/2020\_02\_02, CHN/JS(54)/2020\_03\_27 and CHN/JS(55)/2020\_03\_28) in GenBank and obtained accession numbers: ORF1a, MT415833 ~ MT415839; ORF8, MT415840 ~ MT415846, and downloaded the reference gene sequences from the GISAID (https://www.gisaid.org/), BLAST (NCBI, https://www.ncbi.nlm.nih.gov/) and CGSD websites.

Through analysis of SARS-CoV-2 epidemic strains inside and outside Wuxi city, including the first imported COVID-19 case from Wuhan city in late January 2020 (strain name: CHN/JS(21)/2020\_01\_25, Fig. 2A), 26 local cases directly related to Wuhan, and CHN/JS(13)/2020\_02\_04 strain related to severe outcome (other six strains were related to mild outcome), we performed multiple alignment analyses to identify the subtype of Wuxi prevalent strains. Similar to that observed in most foreign countries, the genotype of Wuxi strains from patients was L-type subtype regardless of severe or mild outcome, and this subtype could be found in many foreign countries as of April 2020 (Fig. 2A). We also found that the S-type was an epidemic strain in Beijing (BJ, January 29, 2020) and Wuhan (WH, January 25, 2020). Homology comparison of the analyzed sequences indicated that the nucleotide identity was 99.7% between the S- and L-type subtype reference strains. The nucleotide and amino acid mutations of SARS-CoV-2 Wuxi strains that appeared in ORF1a (nt 8782, T8517C) were synonymous, whereas the mutation that appeared in ORF8 (nt 28144, C251T, S84L) was nonsynonymous, compared with S-type subtype reference sequences (Fig.2B).

In conclusion, our findings have revealed the clinical features of COVID-19 and the genotypes of SARS-CoV-2 in a regional business hub during the lockdown period, providing a key reference for implementing such study locally and globally to understand SARS-CoV-2 transmission and pathogenesis.

**References**

1. Tian S, Hu N, Lou J, et al. Characteristics of COVID-19 infection in Beijing. *Journal of Infection* 2020;**80**(4):401-06.

2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* 2020;**395**(10223):470-73.

3. Rothe C, Schunk M, Sothmann P, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N Engl J Med* 2020;**382**(10):970-71.

4. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nature medicine* 2020;**26**(4):450-52.

5. Neher RA, Dyrdak R, Druelle V, Hodcroft EB, Albert J. Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. *Swiss medical weekly* 2020;**150**:w20224.

6. Wu AP, Niu PH, Wang LL, et al. Mutations, recombination and insertion in the evolution of 2019-nCoV. *bioRxiv* 2020 Mar 2 (Epub ahead of print).

7. Tang XL, Wu CC, Li X, et al. On the origin and continuing evolution of SARS-CoV-2. *National Science Review* 2020 (Epub ahead of print).

8. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci U S A* 2020;**117**(17):9241-43.

9. Benvenuto D, Giovanetti M, Salemi M, et al. The global spread of 2019-nCoV: a molecular evolutionary analysis. *Pathogens and global health* 2020;**114**(2):64-67.

**Declaration of Competing Interest**

The authors declare no competing interest.

**Acknowledgments**

We thank Xike Zhou for providing technical support.

**Financial support**

This work was funded by the Wuxi Technology Bureau Scientific and Technology Project (CSE31N1607), Wuxi Key Medical Talents Program (ZDRC024), Wuxi Health and Family Planning Commission Appropriate Technology Extension Project (MS201702), National Natural Science Foundation of China (81701550), Emergency Prevention and Control Capacity Program for New Severe Infectious diseases of National Institute for Viral Disease Control and Prevention, the 135 Strategic Program of Chinese Academy of Sciences, and Open Research Fund Program of the State Key Laboratory of Virology of China (2019IOV005).

**Figure Legends**

**Figure 1 Fifty-five SARS-CoV-2 infected cases and migration map. A.** Migration world map of SARS-CoV-2 cases to Wuxi city, Jiangsu Province, China (from January to March 2020). **B.** Geographic migration areas of the SARS-CoV-2 epidemic in Wuxi. Generation 1 was imported cases, and generation 2 and generation 3 cases were found in Wuxi. The number of cases is shown in brackets, while ★ and the number beside it represents the specimen number of sequencing-positive specimen. These positive samples were registered in NCBI GenBank and accession numbers were obtained. **C and D.** Comparison of mild and severe illness case numbers and percentages (mark above the columns) through epidemic generation and sex groups. **E.** Composition comparison with or without comorbid conditions in severe and mild illness cases. Percentages of severe cases or cases with comorbidities are marked above the columns. **F.** Composition percentage of each age group is marked on the pie chart. The number of severe illnesses is marked in brackets below each age composition percentage.

**Figure 2 Maximum likelihood method (MLM) analysis and haplotype analysis of SARS-CoV-2. A.** Phylogenetic analysis of SARS-CoV-2. Analyses of the mutations at key sites in the genome and the phylogenetic tree of the Wuxi strains. PCR primers for SARS-CoV-2 ORF1a and ORF8 were designed referring to the reference strains downloaded from the Coronavirus Global Shared Database website (CGSD, http://nmdc.cn/coronavirus) using Primer Premier 5.0 software, and the primer sequences are as follows: 5’-AAT AAT TGG TTG AAG CAG C-3’ (ORF1a sense) and 5’-TCT ATA AGT TTT GAT GGT-3’ (ORF1a antisense), and 5’-CTT ATT ATC TTT TGG TTC TCC-3’ (ORF8 sense) and 5’-GGG GTC CAT TAT CAG ACA TTT T-3’ (ORF8 antisense), yielding 378 bp and 465 bp DNA products, respectively. Genome sequences were extracted and aligned using BioEdit 7.02 software. Phylogenetic trees for targeted genomes were constructed using Molecular Evolutionary Genetics Analysis (MEGA, *version* 6.06) by the maximum likelihood method (MLM) with 1000 replicates. A black solid circle (●) was plotted before Wuxi strains. The region of the country was expressed in abbreviated form. Each sequence was marked with the accession number and the collection date in the figure. Bootstrap values above 50 are shown. **B.** Haplotype analysis of SARS-CoV-2. Alignment analysis of nucleotide and amino acid sequences was performed with MEGA 6.06. Nucleotide and amino acid sequences of Wuxi strains and reference strains are shown as single letters or dots when they are the same. **Figure 1**



**Figure 2**

H:\202013COVID-2019\gene figure paper\AJE Completed files 6Q7FV87Z\FIG-2.tif