

Mutational Analysis of SARS-CoV-2 Genomes in Key Cities of China

Qiwei Cui¹ and Jeffrey Zheng^{2*}

¹Yunnan University, Kunming, China

²Key Laboratory of Quantum Information of Yunnan, Key Laboratory of Software Engineering of Yunnan, Yunnan University, Kunming, China

***Corresponding Author:** Jeffrey Zheng, Key Laboratory of Quantum Information of Yunnan, Key Laboratory of Software Engineering of Yunnan, Yunnan University, Kunming, China.

Received: August 18, 2020; **Published:** January 29, 2021

Abstract

From December 2019, SARS-CoV-2 induced pneumonia broke out in Wuhan and then spread rapidly from multiple resources to other provinces and other cities in China. In this paper, genomes collected in four Chinese cities: Wuhan, Guangzhou, Shanghai, Hangzhou were analyzed as the A1 module of the MAS. Starting from the virus gene sequence itself, multiple probability statistics are applied to extract characteristics from virus genomes. Variations of genomes can be compared and visualized in such conditions. It is interesting to see various similar and different properties visualized under various groups after transformations. In this way, key mutation characteristics could be observed and this type of results is helpful for further scientific researches on COVID-19 applications.

Keywords: Variation Measurement; Variation Projection; Genome; Visualization; Comparison

Research Background

SARS-CoV-2 is a new type of highly infectious coronavirus. Since the beginning of December 2019, Wuhan City, Hubei Province, China has reported cases of viral pneumonia of unknown causes, which has become the focus of global attention. Chinese scientists quickly identified the novel coronavirus by genetic sequencing and virus isolation. The SARS-CoV-2 belongs to the coronavirus genus, which is closely related to SARS-CoV and MERS-CoV. However, SARS-CoV-2 is not evolved from SARS-CoV and MERS-CoV. Therefore, the problem of virus source has become one of the core issues for scientists in various countries since the beginning of the epidemic. From the perspective of virus gene sequence analysis, scholars in various countries analyze that the source of virus source may come from bats according to the similarity of virus gene sequence, but the exact source needs to be further explored [4] and studied.

At present, the domestic virus outbreak has been effectively controlled, but the virus is spreading rapidly in other areas. European and American countries have become the center of new pneumonia outbreak. SARS-CoV-2 has posed a great threat to the health and safety of people all over the world due to its amazing spreading ability and potential harm.

With the spread of the virus from Wuhan to major cities in China, the gene sequence of the virus itself is constantly changing with the environment changing, so it is necessary to find out the location of virus sequence variation and the degree of variations. So, we focus on the comparison of the gene sequence of the virus and select multiple gene sequences measured in the core cities of China. In view of the sequence variation in different cities, under the framework of variation measurement, we use the A1 model of the MAS, a sequence

Citation: Qiwei Cui and Jeffrey Zheng. "Mutational Analysis of SARS-CoV-2 Genomes in Key Cities of China". EC Neurology SI.02 (2021): 173-178.

comparison model with good effect is constructed, which can accurately locate the variation interval between the sequences of mutated viruses, and determine the variation location for more detailed research and exploration [2].

Aim of the Study

According to the gene sequence of major cities in China, the variation process was analyzed, focusing on Wuhan, Guangzhou, Shanghai and Hangzhou, with the following two focuses:

1. In view of the special case of gene sequence visualization, this paper will focus on gene sequence itself, and explore the distribution law of base number in gene sequence to find the characteristics.
2. In view of the current situation of virus outbreak, a large number of virus gene data are generated every day. For such a large number of data, we can quickly and effectively extract the characteristics and variation information of virus by the way of segmented measurement, which is conducive to more rapid and in-depth research by researchers.

Visualization model of virus sequence

Variation projection model

For the variation projection model of virus sequence, the process is as follows:

- Input: viral gene sequence.
- Processing: segments, measurement.
- Output: Variation projection graph.

In the process of processing the sequence, the total length of the measurement sequence is recorded as N , and the whole sequence is divided into M segments and each segment has m length, then the number of bases contained in each segment is N/M , then M segments are measured and the number of A, G, C, T bases in the sequence is measured in turn, and four groups of bases m_A, m_G, m_C, m_T are obtained, and then the number is projected on the one-dimensional graph.

Variation measurement model

For the variation measurement model of virus sequence, the process is as follows:

- Input: viral gene sequence.
- Processing: segments, measurement, comparison.
- Output: variation measurement graph.

Sum the base numbers of four groups of bases in each section $\sum M_{A,G,C,T}$. Take the absolute value of the difference between the values of the corresponding segments of the two sequences, we can get m values and project them on a one-dimensional graph.

Results and Discussions

According to the transmission time line of the virus, we selected 20 sequences from 4 cities on GISAID for measurement, including 5 in Wuhan, and the detection time of the virus sequence is December 23, 2019; 6 in Hangzhou, testing time is January 22, 2020; 5 in Shanghai, testing time is February 1, 2020; and 4 in Guangzhou, testing time for February 25, 2020 [3].

We can clearly see that the trend of gene sequence visualization is the same in all regions, but we can also see the existence of differences, indicating that the virus has changed in the process of transmission [5].

From the above figure that although the difference can be seen from the projection map, the degree of difference is not obvious. we can compare the Gene sequence between four cities. We can clearly see the location and degree of variation of the virus in the process of transmission.

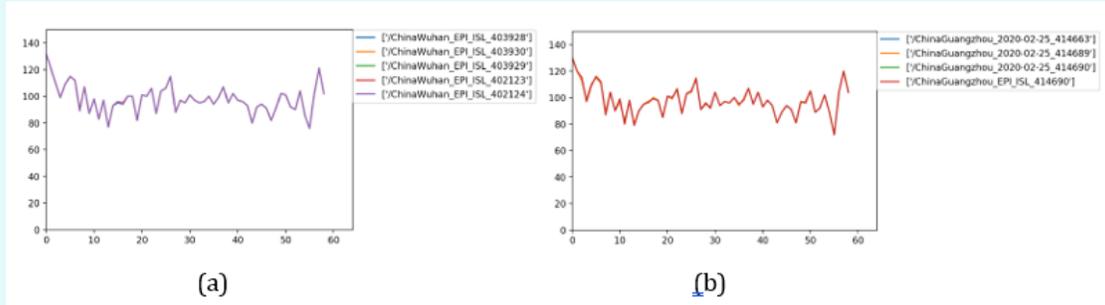


Figure 1: From left to right: variation projection map of Wuhan and Guangzhou.

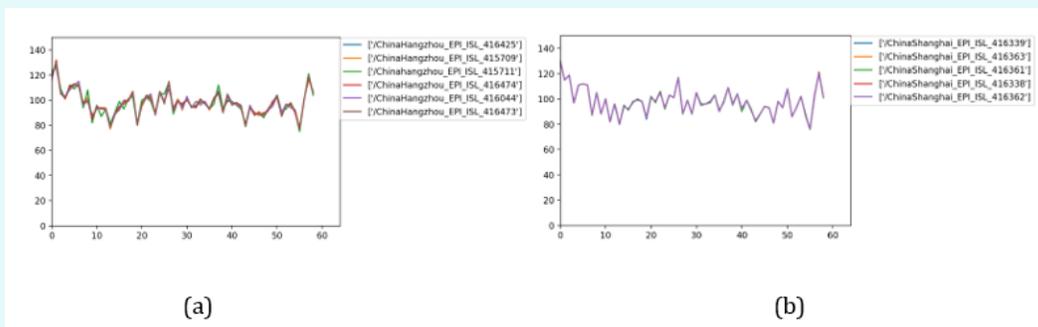


Figure 2: From left to right: variation projection map of Hangzhou and Shanghai.

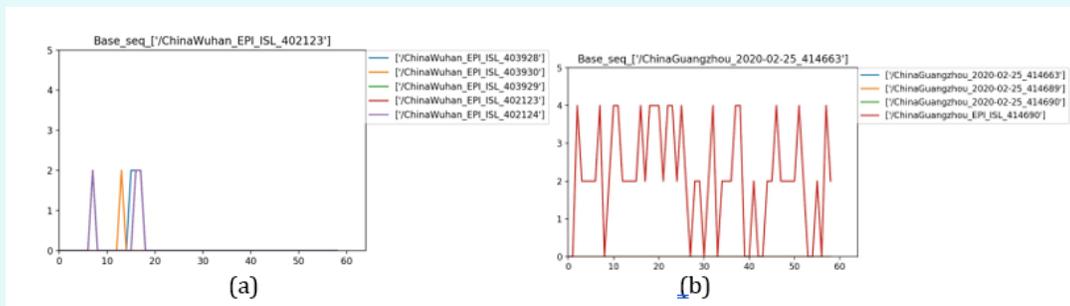


Figure 3: Sequence comparison of multiple genes detected in Wuhan and Guangzhou.

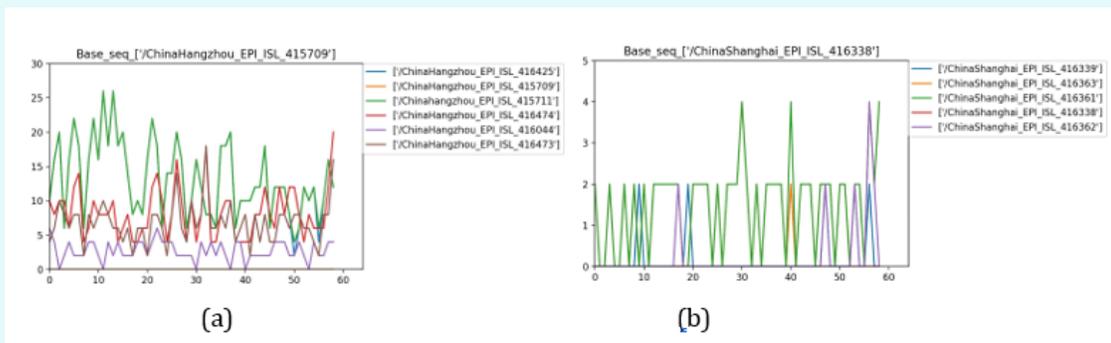


Figure 4: Sequence comparison of multiple genes detected in Hangzhou and Shanghai.

From the above figure, we can see that the similarity of multiple sequences in the same city is different. The number of bases of gene sequences detected from Wuhan, Guangzhou and Shanghai is approximately the same, and the difference between the two sequences is less than 5, which indicates that there is little or almost no variation in the part of the variation when the virus spreads in the city, but the curve in Hangzhou has great changes, which can be inferred that the virus transmitted in Hangzhou may have been transmitted from other areas and changed in the new environment.

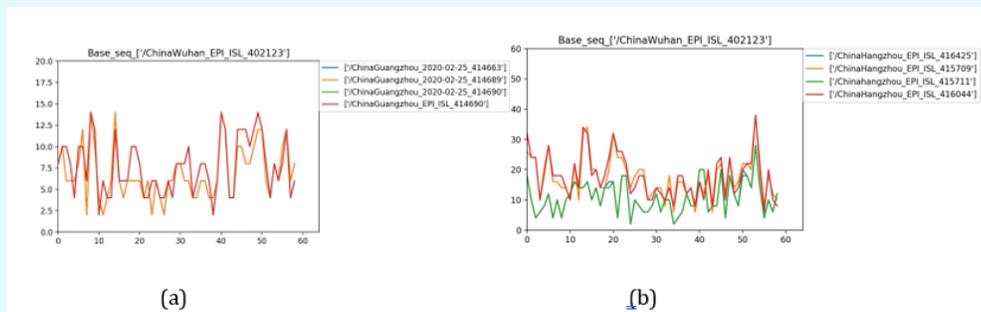


Figure 5: Comparison of gene sequences between Wuhan and Guangzhou, Wuhan and Hangzhou.

We can see the sequence comparison of Wuhan and Guangzhou, Wuhan and Hangzhou, and the change of base difference of corresponding position, which shows the variation of different positions of gene sequence. There are three positions of strong variation points in the comparison diagram of Wuhan and Guangzhou at 8, 14, 40 respectively and three positions of strong variation points in Wuhan and Hangzhou at 14, 20, 54 respectively.

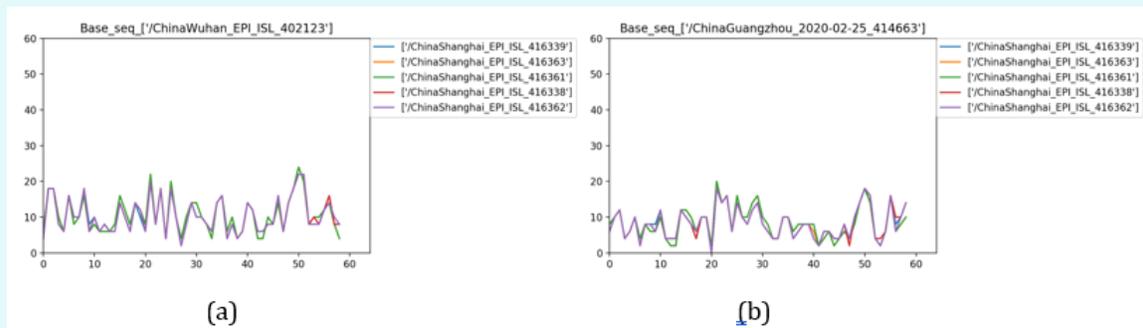


Figure 6: Comparison of gene sequences between Wuhan and Shanghai, Guangzhou and Shanghai.

In comparison of Wuhan and Shanghai, Shanghai and Guangzhou, there are a lot of variation points, and the maximum difference of sequence bases is more than 20, to indicate obvious variations there.

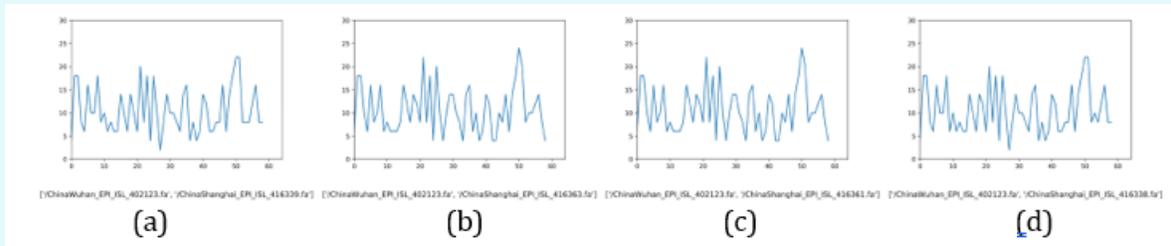


Figure 7: Comparison of gene sequences between Wuhan and Shanghai.

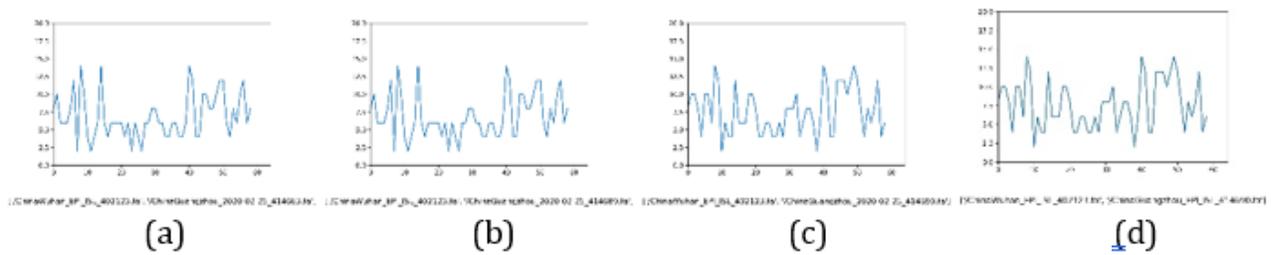


Figure 8: Comparison of gene sequences between Wuhan and Guangzhou.

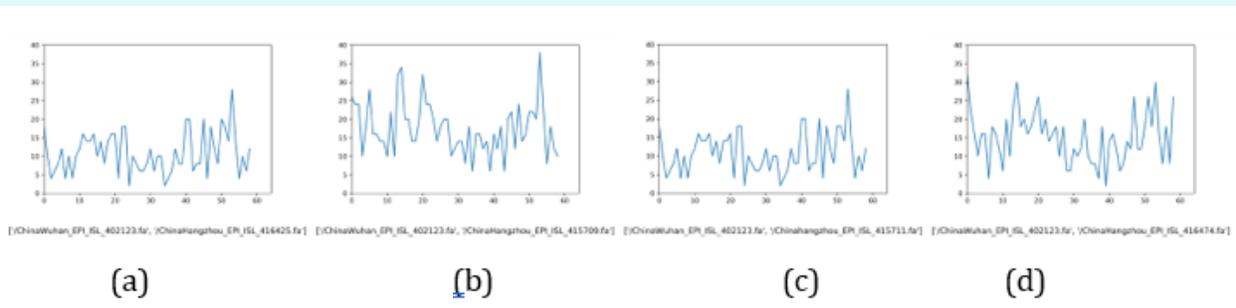


Figure 9: Comparison of gene sequences between Wuhan and Wuhan-Hangzhou.

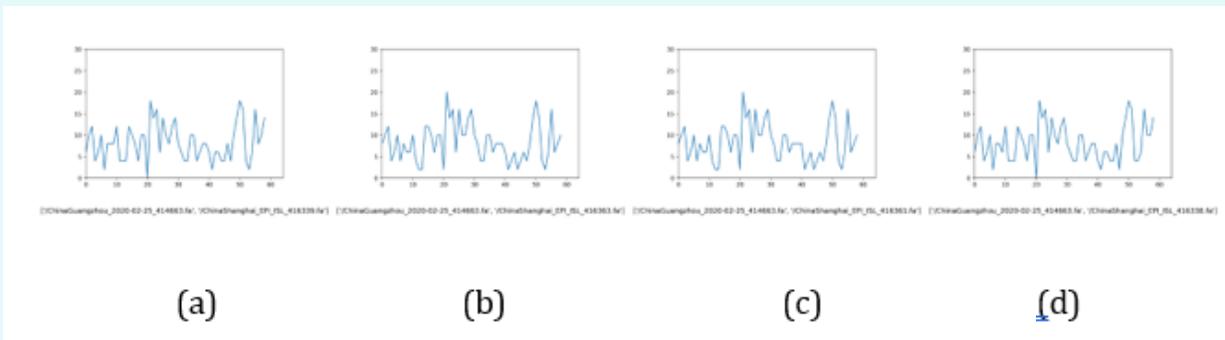


Figure 10: Comparison of gene sequences between Guangzhou and Shanghai.

Conclusion

Under the model of variation measurement, variation projection, for a large number of virus data, various similar and different properties can be visualized under various groups after transformations and in this way we can quickly and clearly identify the characteristics of the virus gene sequence and the position and degree of variation between different virus sequences, variant theory provides a good model and theoretical basis for further research in the future.

Conflict of Interest

No conflict of interest has been claimed.

Acknowledgements

The authors would like to thank NCBI, GISAID, CNGBdb and Next strain for providing the SARS-Cov-2 gene sequence data and other effective information to support this project.

Funding

This work was supported by NSF of China (62041213), the Key Project on Electric Information and Next Generation IT Technology of Yunnan (2018ZI002), NSF of China (61362014), Yunnan Advanced Overseas Scholar Project.

Bibliography

1. Jeffrey Zheng and Chris Zheng. "Biometrics and Knowledge Management Information Systems, Chapter 11: Variant Construction from Theoretical Foundation to Applications". *Springer Nature* (2019).
2. Jeffrey Zheng. "Variant Construction Theory and Applications, Vol. 1: Theoretical Foundation and Applications". Science Press (2020).
3. Qian Xie., *et al.* "Long Research progress of 2019 novel coronavirus". *Journal of Virology* (2020).
4. KaiHao Liang., *et al.* "Analysis of the transmission dynamics of coronavirus". *Journal of University of Electronic Science and Technology* (2020): 27.
5. Thidathip Wongsurawat., *et al.* "Sequences Analysis in thousands of Viral Genomes identify A new common element in Herpesviruses". *Scientific Reports* (2020).
6. Cecilia Saccone and Graziano Pesoe. "Handbook of comparative genomics". Chemical Industry Press (2008).
7. Wu F., *et al.* "A new coronavirus associated with human respiratory disease in China". *Nature* (2020).
8. Jeffrey Zheng. "Variant Construction from Theoretical Foundation to Applications". Springer Nature (2019).
9. Jeffrey Zheng and Chris Zheng. "Variation measurement structure and visual statistical distribution". *Acta Photonica Sinica* (2011).
10. Yuyuan Mao., *et al.* "Mapping Whole DNA Sequence on Variant Maps". 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (2017).
11. Zhihui Hou and Jeffrey Zheng. "Mapping ECG Signals on Variant Maps". 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (2011).
12. Wenjia Liu and Zhijie Zheng. "The Analysis and Visualization System of the Complementary and Matching Features of DNA Sequences". *Hans Journal of Computational Biology* (2015).
13. Yuqian Liu and Jeffrey Zheng. "Visual analysis of encoded and noncoding DNA sequences". *Computational Biology* (2011).
14. Linghua Leng Jeffrey Zheng. "Visualization of ECG Sequence of Sinus Arrhythmia". *Computer Science* (2016).
15. Yuquan Liu and Jeffrey Zheng. "Variant Maps to Identify Coding and Non-coding DNA Sequences of Genomes Selected from Multiple Species". *Biological Systems: Open Access* (2016).

© All rights reserved by Qiwei Cui and Jeffrey Zheng.

Citation: Qiwei Cui and Jeffrey Zheng. "Mutational Analysis of SARS-CoV-2 Genomes in Key Cities of China". *EC Neurology SI.02* (2021): 173-178.