

The role of genomics in COVID-19

Jinrui Wei

Nanjing University of traditional Chinese Medicine, Nanjing 210000, China

Abstract. Since the coronavirus disease 2019 (COVID-19) pandemic swept the globe, the infectious population has proliferated dramatically with each passing day. With the spread of the pandemic around the world, the virus mutated to different lineages in different regions. To better control COVID-19 and trace the source of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2), genomics has played an indispensable role in the epidemic. Scientists have carried out relevant research in genomics and made significant discoveries recently. The author focuses on analyzing the application of genomics to the novel coronavirus pneumonia and delves into the significance of genomics to COVID-19. This paper demonstrates the significance of genomics in tracing the origin of viruses, forecasting the trajectory of the pandemic, and diagnosing and treating COVID-19. This has been of great help in beginning to bring the pandemic under control.

Keywords: COVID-19, Genomics, SARS-COV-2, pandemic.

1. Introduction

Genomics is a subject that studies the composition, structure, expression regulation mechanism and evolution law of genome by means of bioinformatics analysis. Genome research should include two aspects: structural genomics aiming at whole genome sequencing and functional genomics aiming at gene function identification. The research objects are genomic structural characteristics, variation laws and biological significance [1]. It is an interdisciplinary field of science focusing on the structure, function, evolution, mapping, and editing of genomes. A genome is an organism's complete set of DNAs, including all of its genes.

Currently, the mutation of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2) poses a serious threat to public health worldwide. A vast amount of genomics research has been conducted to explore the nature of SARS-COV-2 variants [2, 3], seeking better diagnosis and treatment schemes through the gene sequence information [4]. Scientists are attempting to combat the present situation, and to prevent global outbreaks in the future. With the help of the latest in biotechnology, scientists have been able to unravel the mystery at a remarkable speed. They are reviewing the biology of the new virus more quickly than ever before. And they are hoping it's fast enough to help drive back the disease outbreak.

A recent paper attempts to explore how virulence evolves after a virus jumps to a new host species based on the systematic analysis of the viral genome sequence data [5]. A more comprehensive understanding of virulence mutations and their evolution can be achieved depending on genomics application which is increasingly used to the new virus. The same condition may also apply to searching for the reason why SARA-COV-2 mutated into highly virulent strains, for the control of pathogens and eradication strategy.

2. Applications of genomics in COVID-19

2.1. Trace the origin of SARS-COV-2

As early as January 10, 2020, the genome sequence of SARS-COV-2 was published [6]. Subsequently, relevant genomics research was gradually carried out. Scientists have conducted sequence alignment and evolutionary tree analysis of genome sequences to explore the source and potential host of the virus [7, 8].

In the early stage, researchers found that SARS-COV-2 had the highest sequence homology with bat SARS-like COVs [6, 9, 10]. The virus potentially originates from bats, and there was no trace of

recent recombination, which ruled out the possibility that the virus was produced by artificial recombination [11].

By analyzing coronavirus mutation characteristics, the research team from Sun Yat sen University found that the mutation spectrum of sars-cov-2 was almost the same as that of RaTG13, with 99.9% similarity. This indicates that SARS-COV-2 shared almost the same host environment with RaTG13 before the outbreak [12].

Through comparative genomic analysis, some researchers speculate that SARS-COV-2 may have originated in the recombination of a virus similar to pangolin coronavirus (pangolin-COV) with one similar to a bat coronavirus (RaTG13) [13]. Some researchers suspect that pangolins [14] or minks [15] are intermediate hosts of the virus based on the comparison of different DNA sequences. While some do not endorse that SARS-COV-2 is directly derived from pangolin-COV [16]. At present, both source and intermediate host of SARA-COV-2 remain uncertain. There are ongoing debates among scientists and this area warrants further research to resolve these questions.

However, it is undeniable that genes are the key to making a further recognition of SARS-COV-19. For the first time, researchers found that COVID-19 could be divided into two lineages of "L" and "S" based on two highly linked mutation sites (8782nd and 28144th in the reference genome respectively) and the L-type prevalent in Wuhan may have originated from the S-type [17]. This finding deepens our understanding of this new virus on the genome level and has important reference value for the clinical diagnosis and virus origins.

Since the first full genome sequence of the SARS-COV-2 was shared online, scientists have sequenced and shared about 32,000 viral genomes from all over the world. Such a large amount of data enables researchers to track the source of COVID-19 outbreaks in their country and find out when community transmission occurs. Now, countries that have successfully suppressed the infection are entering the next stage of the COVID-19 with the relaxation of social restrictions, the risk of new cases increases. With the spread of SARS-COV-2 around the world and the gradual evolution of the virus in different regions, different lineages began to form. By comparing sequences, researchers can quickly eliminate possible transmission lines by whether the sequences match or link. Without data from genomics, we can only trace the source of the disease through access. The application of genomics has reduced the investigation pressure of epidemic outbreaks in medical institutions and government departments. Genomics is essential to quickly track and control the outbreak of these diseases. When genomics is used to help contact tracking, the duration of the epidemic tends to be reduced.

2.2. Support virus prevention, control and diagnosis

2.2.1. Guide early prevention and pandemic control

The genome analysis of SARS-COV-2 was carried out in the early stage in China by the researchers from Chinese Academy of Sciences [10]. They found that the genome sequences are almost identical and share 79.6% sequence identity to SARS-COV. The discovery that SARS-COV-2 is highly similar to the known pneumonia virus provides invaluable guidance for early prevention and control [18], especially when it comes to knowing nothing about SARS-COV-2.

In view of this discovery, we can learn from the preventive measures in the SARS period. From past experience and the present situation, we can infer that COVID-19 is mainly transmitted by respiratory droplets. Most of the cases admitted to various hospitals in the early stage had a history of exposure to the South China seafood market in Wuhan. There are also some cases of infection in the family. Inferred from the spread of respiratory diseases, close droplet transmission should be the main way.

Some patients had contact transmission. The disease has a trend of community-wide transmission and spread. Referring to the epidemic prevention measures in the SARS period, we can do a better job in protection and isolation. Most hospitals have established and improved the screening of respiratory infectious diseases, including the process of diagnosis and referral. Every year, as long as it is the season when respiratory infectious diseases are prevalent, all staff will be trained in infectious

diseases. In view of the experience of SARS, we are very cautious in the use of antiviral drugs, hormones and other various drugs this time. This will not lead to drug abuse and plays a good role in the rehabilitation of patients with SARS-COV-2. For individuals, people are supposed to reduce the frequency of going out to avoid infection. If people need to go out in special circumstances, they need to wear disposable medical masks for protection. At the same time, relevant departments should disinfect some areas where the new crown outbreak occurred. For example, specific areas need to be disinfected by special personnel twice a day. Open windows for ventilation for more than half an hour. For the initial epidemic prevention standards and the regional division of hospitals, we can also refer to the SARS period. For example, divide the inpatient area into clean area, potential pollution area and pollution area, and set up a buffer room between two channels and three areas. The boundaries between districts must be clear and the signs must be obvious. The routes of staff and patients need be clear, and all kinds of personnel must follow the specified routes [19] [16].

2.2.2. Predict and Prevent

Whole Genome Sequencing (WGS) is able to read the complete genetic code of SARA-COV-2[20], which helps us to comprehend about the transmission mode and changes of virus and judge whether there are different strains [21]. The sequence data helps solve the real source of the virus exposure of medical staff, proving that they were infected from social activities or hospital patients. Without genomics, we would not know the source of the virus in medical staff through interview. This genetic sequence information prevents the need to investigate possible outbreaks in hospitals. Once we get the sequence, we can infer its source. We can also judge whether the virus has crossed the border and whether the epidemic has further spread. With each additional sequence, researchers are more likely to find the location of the next case.

The government can make use of sequencing to identify the genome, describe the case characteristics and estimate the epidemiological delay distribution. Centers for Disease Control and Prevention can also estimate the doubling time and basic value-added rate of the epidemic in the early stage of exponential growth [22]. By this means relevant departments and institutions can make predictions before the pandemic surges again and take corresponding measures. Public health officials will be able to make faster decisions to control outbreaks. This is of great significance to pandemic prevention.

2.2.3. Diagnose COVID-19

WGS can be used for virus detection as well, but it costs too much both in economy and time. Nowadays reverse transcription-polymerase chain amplification (RT-PCR) has been widely used for the detection of viral RNA [23]. Real-time fluorescent quantitative PCR is regarded as the gold standard for coronavirus detection owing to its high sensitivity, strong specificity, low cost, rapidity and simplicity [24]. It mainly identifies highly conserved sequences in viral RNA sequences. Even if the virus mutates again, these highly conserved sequences remain basically unchanged in evolution [25]. Once the samples are detected positive for COVID-19, the relevant departments can immediately analyze the diversity of genes in the virulent strains by genome sequencing, track the locations of the matched sites in the relevant cases, and exclude the spread of secondary aggregates, providing clues for the transmission chain. This would be of great significance for preventing the further spread of COVID-19.

There is no specific clinical manifestation of COVID-19, which leads to the difficulty to distinguish from common cold and flu [26]. If there was no genomic study on nucleic acid sequence, all patients with suspected symptoms needed to be treated and isolated.

At the same time, genomic data can also enable the government to make a pre judgment "one step ahead" before the beginning of the virus pandemic, so as to take corresponding countermeasures. While this approach also has some limitations. Genome monitoring still relies on extensive nucleic acid detection to capture the corresponding sequences. Some asymptomatic infected people will not take the initiative to accept the virus nucleic acid test. This method of detection can be vulnerable when missed detection occurs.

At present, we can only ensure that asymptomatic infected people are not missed through regular and routine nucleic acid testing and travel monitoring. Real-time fluorescent quantitative PCR can also detect people with asymptomatic infections, which avoids missed diagnosis. If we fail to make use of this technology, we will be unable to diagnose COVID-19 accurately. This will cause the large-scale spread of the epidemic even the irrevocable leeway. Real-time fluorescent quantitative PCR not only provides strong support for clinical diagnosis and discrimination, but also greatly reduces the burden of diagnosis and treatment in hospitals. It is instrumental in monitoring the SARA-COV-2 and responding quickly to the outbreak of the virus, which is greatly conducive to controlling the unprepared pandemic.

2.3. Contribute to drug research and development

Some researchers use the method of single-cell genomics to find the required neutralizing antibody from blood samples of convalescent patients by RNA sequencing of B cells, and prepare high-purity neutralizing antibody to replace plasma as drug to input to patients [27]. Neutralizing antibodies are used to bind to the virus, blocking the virus from invading cells. These neutralizing antibodies can be used to treat COVID-19 and prevent short-term medical care with the intention of protecting the health care workers and patients' families.

Through functional genomics screening, some researchers found a series of host proteins required for virus replication. At the same time, they systematically and comprehensively screened more than 20000 genes in bat cells and identified dozens of key bat genes on which virus replication depends. Finally, it was found that carolacton, an inhibitor of host protein MTHFD1, could effectively inhibit the replication of SARA-COV-2[28] Starting from bat genome analysis, the research used the leading functional genomics methods to systematically search for viral life-cycle dependent host factors. By understanding the molecular mechanism of viral host interaction, they find the new target for antiviral drugs, and then discovered the potential drug molecule carolacton.

The whole gene screening of bats contributed to the discovery of MTHFD1, a new antiviral drug target, and carolacton, an antiviral small molecule. This result also suggests that we can learn how to deal with virus infection from studying the mechanism of virus infection in bats. This will not only help the research and development of COVID-19 drugs and fight the pandemic, but also lay a foundation for mankind to fight against sudden epidemic viruses in the future.

3. Conclusion

Currently, genomics plays an important role in the tracing, diagnosis, prevention and control, and drug development of COVID-19. According to the present outbreak, WGS sequencing of the virus can parse the virus in time, analyze the source of the virus and dynamically monitor the virus variation. To provide a scientific basis for the follow-up study of the impact of virus variation on pathogen detection and vaccine protection effect, and the adjustment of prevention and control policies. COVID-19 is in urgent need of specific therapeutic agents, but there is no effective treatment for the virus. Therefore, it is very important to confirm the drug treatment plan as soon as possible. Through systematic and comprehensive screening of more than 20,000 genes in bat cells, researchers found carolacton, an effective inhibitor of host protein MTHFD1. This discovery will aid in COVID-19's drug research and development and allow us to more effectively fight the current pandemic. It will also lay the foundation for human to fight against the sudden pandemic in the future. The application of genomics enables researchers to find tools to detect SARS-COV-2 and to treat it more effectively.

The application of genomics has also greatly accelerated COVID-19 genome sequencing and analysis, laying the foundation for more comprehensive real-time monitoring of viruses and effective prevention and control of epidemics. The timely use of nucleic acid detection (RT-PCR) can accurately and quickly help the clinicians clarify the cause of fever and promote the accurate clinical diagnosis. Metagenomic sequencing plays a pivotal role in the diagnosis, monitoring, tracking and

tracing of new pathogens, and is a great help for COVID-19 research. It is no exaggeration to say that genomics plays an irreplaceable role in COVID-19.

The traceability, detection, control and drug research and development of COVID-19 remain to be solved. Regarding the diagnosis of asymptomatic infections, better diagnostic methods still need to be found to minimize the threat to public health security. A new study found that covid-19 test based on lateral flow test will be as accurate as laboratory PCR test when it is put into use at the beginning of symptoms [29]. Under the background of gradually relaxing the city closure measures in the future, in addition to continuing to improve the vaccination rate, it will also be very necessary to find mild infections in time through early screening.

In brief, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is still a fast-developing public health crisis. It will be a while before we fully understand its health impact and longer still before there are effective treatments available to fight it. But thanks to the progress in genomics of the coronavirus disease 2019 (COVID-19), we have a greater head-start than ever before.

References

- [1] Li W, and Yin L-P. "Genomics-related concepts and their research progress." *Biology Bulletin*, 2000. 35 (11): p. 3.
- [2] Zhao, X., et al., SARS-CoV-2 Genetic Characterization of Imported from First Case in Ten Local Outbreaks of China. *Chinese Journal of Virology*, 2021. 37 (2): p. 259 - 266.
- [3] Feng, Y., et al., Genetic Characterization of SARS - CoV - 2 from Cross - Border Trucks and Containers in Kashgar Prefecture, Xinjiang Uygur Autonomous Region, China. *Chinese Journal of Virology*, 2021. 37 (6): p. 1292 - 1301.
- [4] Ganna, A., et al., Mapping the human genetic architecture of COVID - 19.
- [5] Geoghegan, J. L. and E. C. J. N. R. G. Holmes, The phylogenomic of evolving virus virulence.
- [6] Wu, F., et al., A new coronavirus associated with human respiratory disease in China. 2020. 579(7798): p. 1 - 8.
- [7] Zhou, P., et al., Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. 2020.
- [8] Implications for the tracking of new coronavirus sequences; Available from: https://www.sohu.com/a/371845478_120380672.
- [9] Dp, A., et al., Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. 79.
- [10] Zhou, P., et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin.
- [11] Chen, L., et al., RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. 2020. 9 (1): p. 313 - 319.
- [12] Deng, S., K. Xing, and X. J. N. S. R. He, Mutation signatures inform the natural host of SARS-CoV-2. 2021.
- [13] Xiao, K., et al., Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. 2020: p. 1 - 7.
- [14] Xiao, K., et al., Isolation and Characterization of 2019-nCoV-like Coronavirus from Malayan Pangolins. 2020.
- [15] Zhu, H., et al., Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm. 2020.
- [16] Liu, P., et al., Are pangolins the intermediate host of the 2019 novel coronavirus (2019-nCoV)? 2020.
- [17] Tang, X., et al., On the origin and continuing evolution of SARS-CoV-2. 2020.
- [18] Zhao Hongjun. Historical experience and practical insights of the Chinese government's response to the impact of the SARS epidemic. *Journal of Hangzhou Normal University: Social Science Edition*, 2020, 42 (2): 8.

- [19] How did the experience of fighting SARS help Peking University People's Hospital to deal with the new crown pneumonia outbreak? Q&A series of the State Council's authoritative conference on joint prevention and control mechanism; Available from: https://www.sohu.com/a/392538386_120578334.
- [20] Chau, D. J. J. o. I., Rapid whole-genome sequencing to inform COVID-19 outbreak response in Vietnam. 2021.
- [21] To, K.W., et al., COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. 2020.
- [22] Pei-Han, L.I., L.I. Peng, and H.B.J.B.B. Song, Application of Metagenomics in Prevention and Control of Infectious Diseases. 2018.
- [23] Chen, X. and S. J. M. i. M. Xia, Sensitive methods for detection of SARS-CoV-2 RNA. 2021.
- [24] Chen L, Xie JG. Interpretation of the "Novel Coronavirus Pneumonia Treatment Protocol (Trial Version 7)"[J]. *Pharmaceutical Guide*, 2020, 39 (5): p. 613 - 615.
- [25] Wu, F., et al., Author Correction: A new coronavirus associated with human respiratory disease in China. 2020. 580 (7803).
- [26] Nichols, L. K., et al., A comparison of coronavirus disease 2019 (COVID-19) versus influenza during the pandemic: Can we distinguish COVID-19 from flu? 2021: p. 1 - 4.
- [27] Chen, C., et al., Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). 2017. 356 (6334): p. 189.
- [28] Anderson, D.E., et al., Orthogonal genome-wide screens of bat cells identify MTHFD1 as a target of broad antiviral therapy. 118 (39): p. e2104759118.
- [29] Wl, A., et al., Comparing the diagnostic accuracy of point-of-care lateral flow antigen testing for SARS-CoV-2 with RT-PCR in primary care (REAP-2). 2021.