

Proximal Origin and Phylogenetic Analysis of COVID-19 (2019-nCoV or SARS-CoV-2)

Attapon Cheepsattayakorn^{1*} and Ruangrong Cheepsattayakorn²

¹10th Zonal Tuberculosis and Chest Disease Center, Chiang Mai, Thailand

²Department of Pathology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

***Corresponding Author:** Dr. Attapon Cheepsattayakorn, 10th Zonal Tuberculosis and Chest Disease Center, Chiang Mai, Thailand.

Received: February 21, 2020; **Published:** March 11, 2020

Abstract

As of February 15, 2020, 51,800 cases of COVID-19 disease, including more than 1,600 COVID-19-related deaths, had been laboratory-confirmed in mainland China, mainly in Hubei province. Additionally, 526 laboratory-confirmed cases have been reported across 25 other countries. Approximately, 15% of cases reported to the World Health Organization (WHO) are severe, 3% are critical, and 82% are mild clinical manifestations, whereas the estimated overall case fatality rate is approximately 2% but the figure outside of Hubei province is approximately 0.05% or less, not different from the fatality identified in the seasonal influenza. It is not yet clear why sustained chains of transmission have not been reported outside Asia. Genetic factors encouraging transmission within Asian populations, effective containment, inefficient transmission, poor reporting due to lacking molecular testing capacity in some low-income countries, and specific environmental conditions in Hubei province and mainland China. Among the most severely affected patients, viral ribonucleic acid (RNA) has been detected in the plasma approximately 15% and viral detection in stool reveals possibility of fecal transmission. The unanswered questions include the pathophysiology of pulmonary clinical infection, influenza and other viral co-infection, and the rate of bacterial complications. Six residuals in the receptor binding domain (RBD) of the spike protein of SARS-CoV and SARS-related coronaviruses, the most variable part of the virus genome appear to be critical for binding to the human ACE2 receptor and the determining host range. Five of these six residuals are mutated in COVID-19 compared to its most closely related virus, RaTG13 sampled from a *Rhinolophus affinis* bat to which it is approximately 96% identical. Volz., *et al.* used Bayesian and maximum likelihood phylogenetic methods on analyzing 53 SARS-CoV-2 (COVID-19) whole genome sequences collected up to February 3, 2020 revealed that the COVID-19 was introduced into the human population in Wuhan, China in early December 2019 and has an epidemic doubling time of about 7 days. In conclusion, to identify the COVID-19 origin, obtaining virus sequences from immediate non-human animal sources would be the most definite method. For substantially refining of phylogenetic estimates of epidemic size and growth rate of COVID-19 in Wuhan, Hubei province and mainland China, larger numbers of more systematically sampled sequences from across China are needed.

Keywords: COVID-19; 2019-nCoV; SARS-CoV-2; Proximal Origin; Phylogenetic

Abbreviations

ACE: Angiotensin-Converting Enzyme; CoV: Coronavirus; COVID-19: Coronavirus Disease 2019; 2019-nCoV: 2019-Novel Coronavirus; MERS: Middle-East Respiratory Syndrome; MRC: British Medical Research Centre; RBD: Receptor Binding Domain; RNA: Ribonucleic

Citation: Dr. Attapon Cheepsattayakorn and Dr. Ruangrong Cheepsattayakorn. "Proximal Origin and Phylogenetic Analysis of COVID-19 (2019-nCoV or SARS-CoV-2)". EC Microbiology SI.02 (2020): 09-12.

Acid; SARS: Severe Acute Respiratory Syndrome; WHO: World Health Organization

Introduction

As of February 15, 2020, 51,800 cases of COVID-19 disease, including more than 1,600 COVID-19-related deaths, had been laboratory-confirmed in mainland China, mainly in Hubei province [1]. Additionally, 526 laboratory-confirmed cases have been reported across 25 other countries [1]. Approximately, 15% of cases reported to the World Health Organization (WHO) are severe, 3% are critical, and 82% are mild clinical manifestations [2], whereas the estimated overall case fatality rate is approximately 2% but the figure outside of Hubei province is approximately 0.05% or less, not different from the fatality identified in the seasonal influenza [3]. There is no expected cross protection by a common human coronavirus infection and theoretically, COVID-19 can infect any one of the individuals [3]. COVID-19 complications target especially the elderly, but the complications can unpredictably occur in the younger age populations, as well as influenza [3]. Despite a low risk of COVID-19 complications, a mild clinical presentations of the disease allow a larger chain of transmission through various populations [3]. Critically, after first identified outbreak, understanding of epidemiology trajectory of infection still have a limitation. Although the WHO reported the dates of diagnosis of COVID-19 disease, but this is not enough information [3]. It is not yet clear why sustained chains of transmission have not been reported outside Asia. Genetic factors encouraging transmission within Asian populations, effective containment, inefficient transmission, poor reporting due to lacking molecular testing capacity in some low-income countries, and specific environmental conditions in Hubei province and mainland China [3]. Among the most severely affected patients, viral ribonucleic acid (RNA) has been detected in the plasma approximately 15% [4] and viral detection in stool reveals possibility of fecal transmission [5]. COVID-19 has been isolated in human saliva [6], nasopharynx and lower respiratory tract [7]. Lacking lung biopsies or post-mortem sample investigations leads to an incomplete understanding of the pathogenesis of COVID-19 infection [3]. The unanswered questions include the pathophysiology of pulmonary clinical infection, influenza and other viral co-infection, and the rate of bacterial complications [3].

Proximal origin of COVID-19

COVID-19 (SARS-CoV-2) is the seventh member of the *Coronaviridae* known to infect humans [8]. Three of these viruses are: 1) SARS CoV-1, 2) MERS and 3) COVID-19 (SARS-CoV-2) can cause severe disease, and 4) HKU1, NL63, OC43 and 229E, are related to mild respiratory symptoms. The two notable characteristics of the COVID-19 genome are: 1) based on structural modelling and early biochemical experiments, COVID-19 is optimized for binding the human ACE2 receptor; and 2) the highly variable spike (S) protein of COVID-19 has a polybasic (furin) cleavage site at the S1 and S2 boundary via the insertion of twelve nucleotides. This event contributes to the acquisition of the three predicted O-linked glycans around the polybasic cleavage site [8].

Six residuals in the receptor binding domain (RBD) of the spike protein of SARS-CoV and SARS-related coronaviruses, the most variable part of the virus genome appear to be critical for binding to the human ACE2 receptor and the determining host range [9]. Five of these six residuals are mutated in COVID-19 compared to its most closely related virus, RaTG13 sampled from a *Rhinolophus affinis* bat to which it is approximately 96% identical [10]. COVID-19 seems to have an RBD that may bind with high affinity to ACE2 from human, non-human primate, cat, pig, and ferret [9]. COVID-19 may bind less efficiently to ACE2 in other species related to SARS-like viruses, such as rodents and civets [9]. Recent binding studies demonstrated that COVID-19 binds with high affinity to human ACE2 [11]. The COVID-19 spike appears to be the result of selection on human or human-like ACE2 permitting another suitable binding solution to occur and this strongly indicate that COVID-19 is not the genetic engineering product [8]. All COVID-19 sequenced genomes have the well adapted RBD and the polybasic cleavage site, and thus are derived from a common ancestor. Initial analyses demonstrated that Malayan pangolins (*Manis javanica*) illegally imported into Guangdong province, China contain a coronavirus (CoV) that is similar to COVID-19 [12,13]. Nevertheless, no pangolin CoV has been identified to be sufficiently similar to COVID-19 across its entire genome for supporting direct human infection [8].

Phylogenetic analysis of COVID-19

Volz, *et al.* demonstrated their study by analyzing 53 SARS-CoV-2 (COVID-19) whole genome sequences collected up to February 3, 2020. They found a strong association between the time of sample collection and accumulation of genetic diversity of COVID-19. By using Bayesian and maximum likelihood phylogenetic methods revealed that the COVID-19 was introduced into the human population in Wuhan, China in early December 2019 and has an epidemic doubling time of about 7 days. Precise estimated of epidemic size are not possible with current genetic data, the analyses demonstrated substantial heterogeneity in the number of secondary infections caused by each COVID-19-infected case that indicated by a high level of over-dispersion in the reproduction number [14].

Conclusion

Although genomic evidence does not support the belief that COVID-19 is a laboratory construct, currently it is impossible to disprove or prove the theories of its origin. To identify the COVID-19 origin, obtaining virus sequences from immediate non-human animal sources would be the most definite method. Additionally, experimental studies of the role of the polybasic cleavage site and predicted O-linked glycans and receptor binding would be more helpful to obtain more viral genetic data. For substantially refining of phylogenetic estimates of epidemic size and growth rate of COVID-19 in Wuhan, Hubei province and mainland China, larger numbers of more systematically sampled sequences from across China are needed.

Author's Contribution

Dr. Attapon Cheepsattayakorn conducted the study framework and wrote the manuscript. Associate Professor Dr. Ruangrong Cheepsattayakorn contributed to scientific content and assistance in manuscript writing. Both authors read and approved the final version of the manuscript.

Competing Interests

The authors declare that they have no actual or potential competing financial interests.

Funding Sources

The authors disclose no funding sources.

Bibliography

1. World Health Organization. Coronavirus 2019 (COVID-19): situation report-26, February 15, 2020. Geneva: World Health Organization (2020).
2. WHO press conference (2020).
3. Vetter P, *et al.* "Covid-19: a puzzle with many missing pieces, with better information on epidemiology, pathogenesis, and treatments are urgent priorities". *British Medical Journal* 368 (2020): m627.
4. Huang C., *et al.* "Clinical features of patients infected with 2019 novel coronavirus in Wuhan". *Lancet* 395.10223 (2020): 497-506.
5. Zhang H., *et al.* "The digestive system is a potential route of 2019- nCoV infection: a bioinformatics analysis based on single-cell transcriptomes". *BioRxiv* (2020).
6. To KK., *et al.* "Consistent detection of 2019 novel coronavirus in saliva". *Clinical Infectious Diseases* (2020).
7. Zhu N., *et al.* "China novel coronavirus investigating and research team. A novel coronavirus from patients with pneumonia in China, 2019". *New England Journal of Medicine* 382.8 (2020): 727-733.

8. Andersen KG, *et al.* "The proximal origin of SARS-CoV-2". Arambaut ARTIC Network (2020): 9.
9. Wan Y, *et al.* "Receptor recognition by novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS". *Journal of Virology* (2020).
10. Wu F, *et al.* "A new coronavirus associated with human respiratory disease in China". *Nature* (2020).
11. Wrapp D, *et al.* "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation". *bioRxiv* (2020).
12. Liu P, *et al.* "Viral metagenomics revealed sendai virus and coronavirus infection of Malayan pangolins (*Manis javanica*)". *Viruses* 11.11 (2019): 979.
13. Wong MC, *et al.* "Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019". *bioRxiv* (2020).
14. Volz E, *et al.* "WHO Collaborating Centre for Infectious Disease Modelling, MRC Centre for Global Infectious Disease Analysis, Adult Latif Jameel Institute for Disease and Emergency Analytics (J-IDEA), and Imperial College London. Report 5: Phylogenetic analysis of SARS-CoV-2" (2020): 7.

© All rights reserved by Dr. Attapon Cheepsattayakorn and Dr. Ruangrong Cheepsattayakorn.