SEQUENCE ANALYSIS AND PHYLOGENETIC RELATIONSHIP OF COVID-19-WUHAN-HU-1

Farooqe E Azam[a],[*], Nausheen Saba Khanam[b], Aabsheen Saba Khanam[c]

[a] Department of Applied Science, J.S. University, Shikohabad, Uttar Pradesh, India
[b] Department of Zoology, School of Life Sciences, Dr. B.R. Ambedkar University, Agra, India
[C] Department of Chemistry, Agra College, Dr. B.R. Ambedkar University, Agra, India
[*]Corresponding author: farooqejsu@gmail.com

ABSTRACT

A novel coronavirus associated with human-to-human transmission and severe human infection has been reported from Wuhan in China and spread worldwide. Our objectives were to characterize the genetic relationships of COVID-19-Wuhan-Hu-1 and some other coronaviruses. Viral diseases are widely spread infections caused by viruses, a type of microorganism. Viruses can cause a wide range of diseases and come in a variety of shapes and sizes. The most common type of viral illness is the common cold, which is caused by a viral infection of the upper respiratory tract. Bats have long been known to be natural reservoirs for a wide range of viruses. Bat coronaviruses have received special attention as one of the new coronaviruses that have produced unexpected human illness outbreaks around the world. COVID-19-Wuhan-Hu-1 was suggested to have originated from bats. Here we are going to find the phylogenetic relationship of COVID-19-Wuhan-Hu-1 with Pipistrellus Bat Coronavirus-HKU5, Human Coronavirus-NL63, Porcine Coronavirus-HKU15, Avian Infectious Bronchitisvirus-I, Beluga Whale Coronavirus-SW1, and Middle East Respiratory Syndrome Coronavirus. Evolutionary relationships are frequently described by phylogenetic trees, which are the most useful for genetic studies. We used a percent identity metric-based method for comparing trees, which extracts distinct alternative evolutionary relationships. Understanding the phylogenetic relationship of COVID-19-Wuhan-Hu-1 with other viruses is helpful for the prediction and prevention of another pandemic emergence in the future.

Keywords: ClustalW, COVID-19, Evolutionary relationship, Phylogenetic tree, Sequence analysis

1. Introduction

A new coronavirus (COVID-19-Wuhan-Hu-1) has severe human infection, which was reported from the city of Wuhan, China in 2019. The coronavirus has become the major pathogen behind emerging respiratory disease outbreaks. They are a large family of single-stranded RNA viruses that can be isolated in different animal species [1]. These viruses can cross species barriers and can cause illnesses in humans ranging from the common cold to more severe diseases such as COVID-19. Symptoms of viral infections vary in nature, quality, and severity based on the type of viral infection as well as other factors such as the person's age and overall health. Flu-like symptoms and a chilly fever are common symptoms of viral infections. Viral diseases are not treatable with antibiotics, which can only cure bacterial and fungal diseases and infections [2]. However, the most common

viral diseases, such as the common cold, respiratory infections, and the flu, are self-limiting in generally healthy people. This means that the viral infection causes illness for a period of time and weakens the immune system of the body.

The family Coronaviridae includes a large number of viruses that in nature are found in bats, humans, and birds. Human coronavirus, first characterized in the 1960s, is associated with a large percentage of respiratory infections. Scientific interest in coronaviruses exponentially increased after the emergence of SARS-Cov in Southern China [3]. The virus was initially detected in Himalayan palm civets, which may have served as an amplification host. The civet virus contained a 29-nucleotide sequence generally not found in most human isolates that were related to the global epidemic. The function of the affected Open Reading Frame-10 might have played a role in the trans-species jump and a similar virus was found later in bats. After the SARS epidemic, bats have been considered as a potential reservoir species that could be implicated in future coronavirus-related human pandemics [4]. The Middle East Respiratory Coronavirus (MERS-CoV) first appeared in Saudi Arabia in 2012. Camels are thought to have played a key role in the virus's transmission to humans, and the virus's genesis has been traced back to bats. Coronavirus probably originated in bats and then moved into other mammalian hosts, such as the Himalayan palm civet for SARS-Cov-2, before transferring to humans. Studies in species other than COVID-19-Wuhan-Hu-1 have examined hostvirus phylogeny and identified coevolutionary relationships or incongruous phylogenetic patterns. Studies have demonstrated that the relationship between viral phylogeny and geographic location and identification hosts can produce information on the origin of the coronavirus. We know the high genetic diversity of bat coronaviruses and carried out a systematic phylogenetic study of the viruses and their hosts to examine evolutionary relationships between coronaviruses and bats [5]. The aim was to further investigate the origins of COVID-19-Wuhan-Hu-1 and SARS. We discuss host viral traits that might have given rise to these patterns and comment on the implications of our findings for the emergence of COVID-19-Wuhan-Hu-1.

In the study of evolutionary relationships, a given set of organisms that have different phylogenetic characteristics can be inferred from each combination of input data. A phylogenetic tree provides information about different organisms that have some relationship with each other. So the individual gene trees carry different evolutionary signals [6]. The phylogenetic tree provides useful ideas about evolutionary relations on the basis of genetic information. These visualizations can be easy to interpret when there are some species to be compared. Quantitative metric-based tree comparisons are an alternative to visual methods, which can be useful for evolutionary studies. Here we used a percent identity matrix-based method for the study of phylogenetic relations of coronaviruses. A metric is a mathematical concept of distance; specifying a metric gives structure and shape to a set of objects. Each metric on a set of trees defines the tree space [7]. The size and complexity of tree spaces present serious challenges to possible topologies for rooted and binary trees. Here we present an approach to comparing and clustering groups of coronavirus species in a phylogenetic tree. A tree metric that allows for unambiguous low-dimensional visualisations of tree space. It allows for the easy identification of distinct groupings of similar trees. This method is a logical solution to the challenge of summing up complicated tree spaces, resulting in a small number of representative trees with different evolutionary patterns reflected in the data [8].

2. Material and methods

2.1 Sequences of Genomes

All the coronavirus sequences were downloaded from the NCBI nucleotide sequence database (http://www.ncbi.nlm.nih.gov). Only coronaviruses were included in this study. We used full genome sequences that were obtained from seven coronavirus species (COVID-19-Wuhan-Hu-1 0.24688, Pipistrellus Bat Coronavirus-HKU5 0.26617, Human Coronavirus-NL63 0.25462, Porcine Coronavirus-HKU15 0.31909, Avian Infectious Bronchitisvirus-I 0.22144, Beluga Whale Coronavirus-SW1 0.14207, Middle East Respiratory Syndrome Cor 0.23242).

2.2 Alignment of multiple sequences

After the collection of all the seven full genome sequences, here we need to align. For alignment of more than two gene sequences, the clustal package provided an online facility for aligning the multiple sequences. In Figure 1, the multiple genomic sequence alignment was performed using CustalW2 (https://www.ebi.ac.uk/tools/clustalw) and manually edited using MEGA v1.0 according to the encoded reading frame [9]. The percent identity matrix was calculated by the Clustal 2.1 for identical relations of gene sequences shown in figure 3.

2.3 The phylogenetic tree

A phylogenetic tree was constructed using the MUSCLE 3.8 (https://www.ebi.ac.uk/Tools/muscle). The phylogenetic clustering of the full-length genome was constructed using the neighbor-joining method without distance corrections [10]. The phylogeny.fr is used to display various coronavirus sequences for evolutionary relationships within the COVID-19-Wuhan-Hu-1 genome. The visualization of phylogenetic trees in the form of cladograms Phylogenetic trees were viewed using FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree).

```
PipistrellusBatCoronavirus-HKU5     ATGA------------------------------------------------------
HumanCoronavirus-NL63               ATGA------------------------------------------------------
PorcineCoronavirus-HKU15            ATGGCC---------------------------------AAGAACAAGTCCAAGCGC
COVID-19-Wuhan-Hu-1                 ATGT------------------------------------------------------
AvianInfectiousBronchitisvirus-I    ATGT------------------------------------------------------
BelugaWhaleCoronavirus-SW1          ATGA------------------------------------------------------
MiddleEastRespiratorySyndromeCor    ATGTCTTTCGTGGCTGGTGTGACCGCGCAAGGTGCGCGCGGTACGTATCGAGCAGCGCTC
                                    ***


PipistrellusBatCoronavirus-HKU5     ------------------------------------------------------------
HumanCoronavirus-NL63               -AACTTTTCTTGATTTTGCTTGTTTTTGCCCCTGGCCTCTTGCTTTTTCACATGTAATAGT
PorcineCoronavirus-HKU15            GACGCTATTGCGTTGCCTGAAAATGTACCACCACCTCTGCAACTTTTCATTCATGTTGCA
COVID-19-Wuhan-Hu-1                 ------------------------------------------------------------
AvianInfectiousBronchitisvirus-I    ------------------------------------------------------------
BelugaWhaleCoronavirus-SW1          ------------------------------------------------------------
MiddleEastRespiratorySyndromeCor    AACTCTGAAAAACATCAAGACCATGTGTCTCTAACTGTGCCACTCTGTGGTTCAGGAAAC


PipistrellusBatCoronavirus-HKU5     --------------------------------------------------------TACGC
HumanCoronavirus-NL63               AATGCTAATCTCTCTATGTTACAATTAGGTGTTCCTGACAATTCTTCAACTATTGTTACG
PorcineCoronavirus-HKU15            GCTGCTGAAGAGGGTCACCCTAAGGTTACTACT--------TACCTTGGCAACTATAACC
COVID-19-Wuhan-Hu-1                 ----------------------------------------------------TTGTTTTT
AvianInfectiousBronchitisvirus-I    ----------------------------------------------------TGGTAACA
BelugaWhaleCoronavirus-SW1          ------------------------------------------------------------
MiddleEastRespiratorySyndromeCor    CTGGTTGAAAAACTTTCACCATGGTTCATGGATGGCGAAAATGCCTATGAAGTGGTGAAG


PipistrellusBatCoronavirus-HKU5     TCAGTGTTAGT----------------ACTGATGTGCTCGTTAACTTTTATAGGAAACC-
HumanCoronavirus-NL63               GGTTTATTGCC--------AACTCATTGGTTTTGTGCTAATCAGAGTACATCTGTTTAC-
PorcineCoronavirus-HKU15            TCTATGCCACC-------AAGGCTCCGCCTGGCGTGCAGGTTCTTAGTGCTAAAACCTC-
COVID-19-Wuhan-Hu-1                 CTTGTTTTATT--GCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAACCAGAACTCAA-
AvianInfectiousBronchitisvirus-I    CCTCTTTTACT-------AGTGACTCTTTTGTGTGCAC--------------------
BelugaWhaleCoronavirus-SW1          ------------------------------------------------------------
MiddleEastRespiratorySyndromeCor    GCCATGTTACTTAAAAAGGAGCCACTTCTCTATGTGCCCATCCGGCTGGCTGGACACACT
```

Figure 1 ClustalW2-MUSCLE genome sequence alignment of multiple coronavirus sequences

3. Result and discussion

In our study, the noval coronavirus (COVID-19-Wuhan-Hu-1) is not a mosaic and it is most closely related to the Pipistrellus Bat Coronavirus-HKU5 detected in the bat [11]. The levels of genetic similarity between COVID-19-Wuhan-Hu-1 and Pipistrellus Bat Coronavirus-HKU5 indicate that the latter does not provide the exact variant that caused the human outbreak, but the hypothesis that COVID-19-Wuhan-Hu-1 originated from bats is very likely [12]. On the other hand, there is evidence for discordant phylogenetic relationships between COVID-19-Wuhan-Hu-1 and the Pipistrellus Bat Coronavirus-HKU5 clade with their closest partners, shown in figure 2. According to a recent study based on codon use analysis, the spike protein of COVID-19-Wuhan-HU-1 may have evolved through recombination from an unknown unsampled coronavirus. The recently published bat coronavirus sequence gives sufficient phylogenetic information to determine the origin of the Spike protein as well as the remainder of the genome, implying that the genome has a common ancestor. COVID-19-Wuhan-genetic Hu-1's traits and their potential association with virus properties and pathogenicity in humans are still unknown.
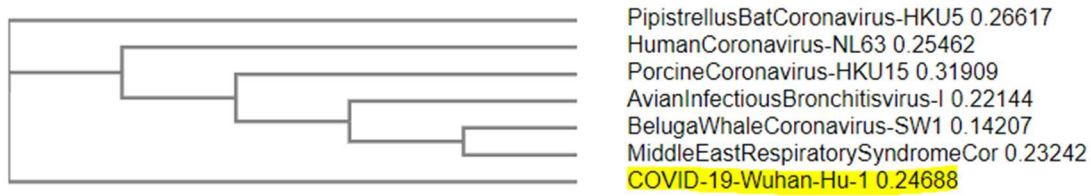
A ES

PipistrellusBatCoronavirus-HKU5 0.26617
HumanCoronavirus-NL63 0.25462
PorcineCoronavirus-HKU15 0.31909
AvianInfectiousBronchitisvirus-I 0.22144
BelugaWhaleCoronavirus-SW1 0.14207
MiddleEastRespiratorySyndromeCor 0.23242
COVID-19-Wuhan-Hu-1 0.24688

Figure 2 Phylogenetic tree showing evolutionary relationship of seven coronavirus gene sequences by

Figtree v1.4 tool

```
1: PipistrellusBatCoronavirus-HKU5      100.00   47.64   40.35   48.70   48.49   54.26   40.81
2: HumanCoronavirus-NL63                 47.64  100.00   41.80   49.22   50.65   54.55   44.70
3: PorcineCoronavirus-HKU15              40.35   41.80  100.00   42.04   43.32   47.49   43.73
4: COVID-19-Wuhan-Hu-1                   48.70   49.22   42.04  100.00   51.70   55.18   44.97
5: AvianInfectiousBronchitisvirus-I      48.49   50.65   43.32   51.70  100.00   59.34   51.43
6: BelugaWhaleCoronavirus-SW1            54.26   54.55   47.49   55.18   59.34  100.00   62.55
7: MiddleEastRespiratorySyndromeCor      40.81   44.70   43.73   44.97   51.43   62.55  100.00
```

Figure 3 Percent identity matrix of coronavirus genome sequences
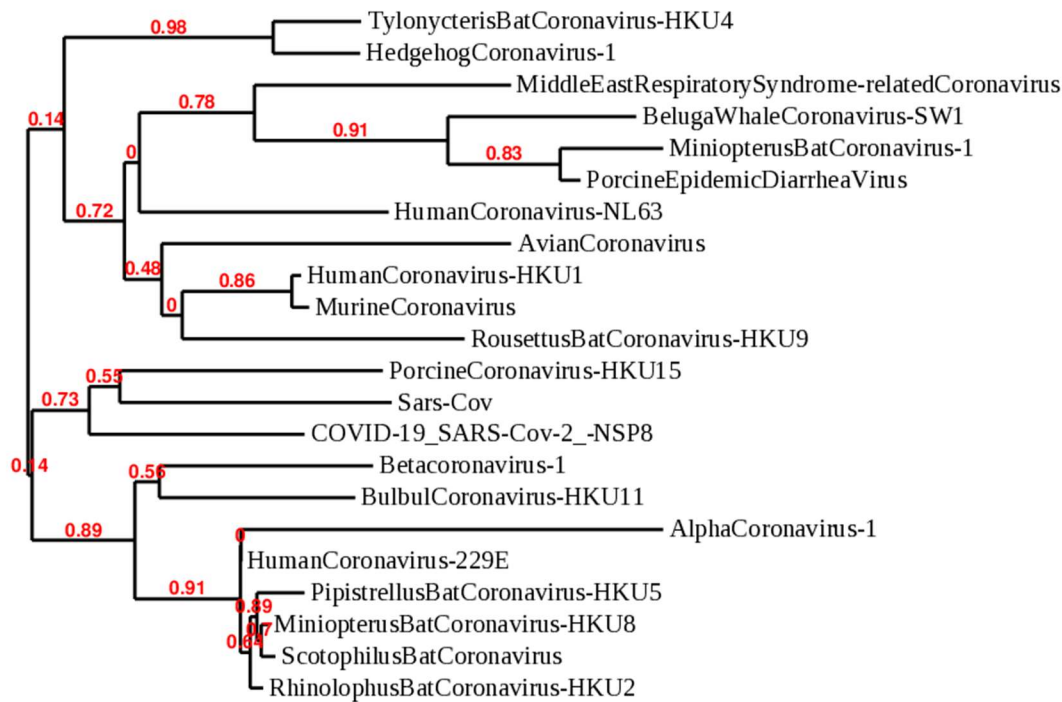
Figure 4 Phylogenetic tree showing relationship of various coronavirus gene sequences by phylogeny.fr

On the basis of the given phylogram shown in Figure 4, the evolutionary relationship of COVID-19-Wuhan-Hu-1 = 0.73, Pipistrellus Bat Coronavirus-HKU5 = 0.98, Human Coronavirus-NL63 = 0.78, Porcine Coronavirus-HKU15 = 0.55, Avian Infectious Bronchitis-virus-I = 0.48, Beluga Whale Coronavirus-SW1 = 0.91, Middle East Respiratory Syndrome Cor = 0.78. By combining information

derived from the phylogram of coronaviruses with data on the geographic origin of viruses, we were able to describe the phylogeographic distributions for known coronaviruses from COVID-19-Wuhan-Hu-1. When we mapped a cladogram of various species, the results showed genome-based evolutionary relations. Here, the phylogenetic relationships between virus lineages were similar across our analyses, and well-supported genetic structure was observed within some coronaviruses. Thus, evolutionary patterns may reflect either a high frequency of host shifts among closely related bat coronaviruses

4. Conclusion

In the conclusion of our study, an outbreak of novel coronavirus was detected in Wuhan in the winter of 2019–20. Wuhan, like Guangzhou during the 2003 SARS pandemic, is a fast growing capital city in Hubei province and a major traffic hub in central China. Furthermore, both outbreaks were linked to wet marketplaces where game animals and meat were marketed at the time. Unlike the 2003 SARS outbreak, which began as an animal-to-human transmission during the early stages of the epidemic, China's improved surveillance network and laboratory capability were able to recognise this outbreak within a few weeks and announce the virus genome sequences, allowing the development of rapid diagnostic tests and efficient epidemiological control. We can discover the trend of genetic diversity through phylogenetic analyses. In the evolution of various species, the identical relationship is very significant. Bats are becoming more widely recognised as reservoirs for a wide range of highly fatal zoonotic diseases. Understanding their diversity, behaviour, and virus transmission methods could help prevent future outbreaks of known and new zoonotic illnesses spread by bats.

5. Acknowledgements

6. References

[1] Beck, B.R., Shin, B., Choi, Y., Park, S., Kang, K., 2020. Antiviral drugs that may act on   the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput. Struct. Biotechnol. J. 18, 784–790. https://doi.org/10.1016/j.csbj.2020.03.025.

[2] Chan, J.F.W., Kok, K.H., Zhu, Z., Chu, H., To, K.K.W., Yuan, S., Yuen, K.Y., 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with a typical pneumonia after visiting Wuhan. Emerg. Microbes Infect. 9, 221–236. https://doi.org/10.1080/22221751.2020.1719902.

[3] Ginestet, C., 2011. ggplot2: Elegant graphics for data analysis. J. R. Stat. Soc.: Ser. A (Stat. Soc.) 174, 245–246. https://doi.org/10.1111/j.1467-985X.2010.00676_9.x.

[4] Hegyi, A., Ziebuhr, J., 2002. Conservation of substrate specificities among coronavirus main proteases. J. Gen. Virol. 83, 595–599. https://doi.org/10.1099/0022-1317-83- 3-595.

[5]   Holmes EC. The phylogeography of human viruses. Mol Ecol. 2004;13:745–56.

[6]   Ithete, N.L., et al., 2013. Close relative of human MERS coronavirus in bat, South Africa. Emerg. Infect. Dis. 19 (10), 1697–1699.

[7] Joshi, R.S., Jagdale, S.S., Bansode, S.B., Shankar, S.S., Tellis, M.B., Pandya, V.K., Chugh, A., Giri, A.P., Kulkarni, M.J., 2020. Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease. J. Biomol. Struct. Dyn. 1–16. https://doi.org/10.1080/ 07391102.2020.1760137.

[8]   Kahn, J.S., McIntosh, K., 2005. History and recent advances in coronavirus studies. Pediatr. Infect. Dis. J. 24 (11 Suppl) p. S223–7, discussion S226.

[9]   Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA v1.0: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096.

[10]   Li W, Shi Z, Yu M, Ren W, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005;310: 676–9.

[11]   Luk, H.K.H., Li, X., Fung, J., Lau, S.K.P., Woo, P.C.Y., 2019. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. Infect. Genet. Evol. 71, 21–30. https://doi.org/10.1016/j.meegid.2019.03.001.

[12]   Traditional Chinese Medicine Database (TCMD), 2009. Beijing.