J Pharm Biopharm Res, 2020, 2(1): 126-130
DOI: 10.25082/JPBR.2020.01.004

RESEARCH ARTICLE

Phylogenetic analyses and genomic variation of the 2019-nCoV

Faiz Ul Haq[1,2*] Sidrah Saleem[1] Muhammad Imran[1] Ayesha Ghazal[1] Kashif Ahmad[2,5] Muhammad Roman[1] Saeed Ur Rahman[3] Sami Ullah[4] Iftekhar Ahmad[2] Habibah Mehmood[4] Wajahat Ullah[2]

Abstract: There is a rising global concern about the SARS CoV-2 as a public health threat. Complete genome sequence have been released by the worldwide scientific community for understanding the molecular characteristics and evolutionary origin of this virus. Aim of the current context is to present phylogenetic relationship and genomic variation of 2019-nCoV. Based on availability of genomic information, we constructed a phylogenetic tree including also representatives of other *coronaviridae*, such as Middle East respiratory syndrome, severe acute respiratory syndrome and Bat coronavirus. The phylogenetic tree analysis suggested that SARS CoV-2 significantly clustered with bat SARS like coronavirus genome, however structural analysis revealed mutation in Spike Glycoprotein and nucleocapsid protein. However our phylogenetic and genomic analysis suggests that bats can be the reservoir for this virus. Lack of forest might be the fact in association of bats with human environment. It is also difficult to study on bats due to absence of proper reagent and availability of few species for research. We confirm high sequence similarity ($> 99\%$) among sequenced SARS CoV-2 genomes, and 96% genome identity with the bat coronavirus, confirming the notion of a zoonotic origin of SARS CoV-2.

Keywords: SARS-CoV-2, coronavirus, CLUSTAL analysis, genomic variation

1 Introduction

Coronaviridae is a family of the largest known viruses with single stranded RNA genome[1]. Based on phylogenetic analyses and antigenic criteria, They have been categorized in three groups[2], specifically: (1) alpha CoVs, responsible for gastrointestinal disorders in human, cats, pigs, and dogs; (2) beta CoVs, including the human severe acute respiratory syndrome (SARS) virus, Middle Eastern respiratory syndrome (MERS) virus and Bat coronavirus (BCoV); (3) gamma CoVs, which responsible for infection in avian species.

Very recently, a novel beta-CoVs coronavirus (2019-nCoV) has been causally linked to severe respiratory infection in humans. Phylogenetic relationships between Human and Bat *coronaviridae* have been discovered for SARS[3] and more recently also for 2019-nCoV, Benvenuto *et al.*[4] suggested the hapening of inter species transmission[5]. Later the novel coronavirus-2019 was designated as Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) and the disease is called COVID-19[6,7].

No vaccine for SARS cov-2 has been released up to yet, but a World effort and trials has ascended toward the characterization of the molecular determinants and evolutionary features of SARS cov-2 vaccine[8]. A preliminary assessment of 10 genomic sequences from 2019-nCoV specimens has described a low heterogeneity of this viruses with inter-sample genome identity above 99.9%[5].

Although several studies have published phylogenetic trees of the SARS cov-2, inwhich some of the analysis including fragment sequences but do not provide accurate phylogenetic information. Therefore, we analyzed the available viral genome sequences and retrieve only full length viral genomic sequences for phylogenetic analysis. As well as we set out to characterize the heterogeneity of SARS CoV-2 genomes and proteomes, and comparing them with other related viruses of *coronaviridae*, specifically, Bat corona virus, MERS-Cov, Pangolin-Cov and SARS-Cov.

2 Methods

Reference virus genomes were retrieve on 16 March 2020 from GenBank using Blastn with accession num-

Received: June 14, 2020 **Accepted:** August 12, 2020 **Published:** August 24, 2020

*Correspondence to: Faiz Ul Haq, Department of Microbiology University of Health Sciences Lahore, Pakistan; Email: faizulhaq553@gmail.com
[1] Department of Microbiology University of Health Sciences Lahore, Pakistan
[2] Center for Biotechnology and Microbiology University of Swat, Pakistan
[3] Department of Nursing University of Health Sciences Lahore, Pakistan
[4] Department of Forensic Science University of Health Sciences Lahore, Pakistan
[5] Department of Microbiology, Hazara University, Pakistan

Citation: Haq FU, Saleem S, Imran M, *et al*. Phylogenetic analyses and genomic variation of the 2019-nCoV. *J Pharm Biopharm Res*, 2020, **2**(1): 126-130.

ber (AY274119.3) as a query. DNA sequence alignment were done using CLUSTAL X with a pairwise multiple alignment parameter[9]. Nucleotide sequence coverage and identity were intended using BLAST nucleotide v2.6.0[10].

Phylogenetic model generation and tree visualization were done using software MEGA7 v 7.0.26[11] with the Maximum Likelihood method[12]. The structure of tree was validated by running the analysis on 100 bootstrapped datasets input[13].

The Bayesian analysis was accompanied in MrBayes 3.1 under the GTR + I + G model[14]. The Markov chain Monte Carlo chains (MCMC) were run simultaneously with generations of 1000000 and resulted trees were sampled every hundred generations. The first 1000 trees were discarded (burn-in) and the remaining trees were used to construct a model tree with posterior probability distribution. We used recently developed S-DIVA and BBM analyses implemented in RASP to reconstruct the possible ancestral ranges of family *Coronaviridae* on the phylogenetic trees[15].

Algorithms of every node were calculated using software RASP generated by S-DIVA and BBM analysis. The S-DIVA algorithm In RASP encode four types of biogeographic events namely, vicariance, dispersal, extinction and duplication. Bayesian Binary MCMC (BBM) and S-DIVA analysis were run in RASP that show us possible vicaraince and dispersal event with probability algorithms[16].

3 Results

Phylogenetic analysis and genomic variation was conducted using molecular data from NCBI for the assessment of evolutionary relationship among 2019-nCoV sampled from various geographical location.

The gene bank information with accession number

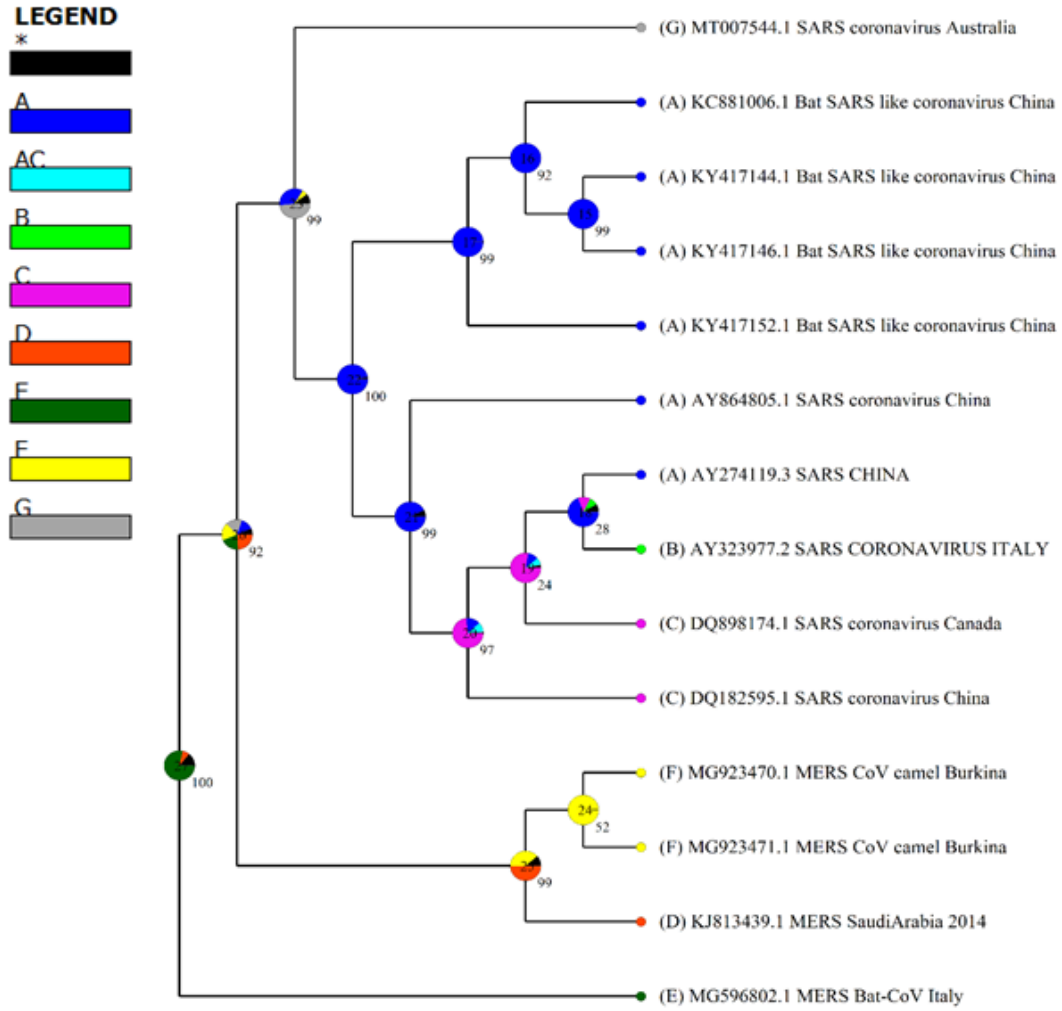

Figure 1. Phylogenetic tree of 2019-novel Coronavirus sequences, Bat coronavirus and Middle Eastern respiratory syndrome. Colour key indicates biogeographical regions: A, China (SARS); B, Italy (SARS); C, Canada (SARS); D, Saudi Arabia (MERS); E, Italy (Bat-CoV); F, Burkina (Camel); G, Australia (SARS)

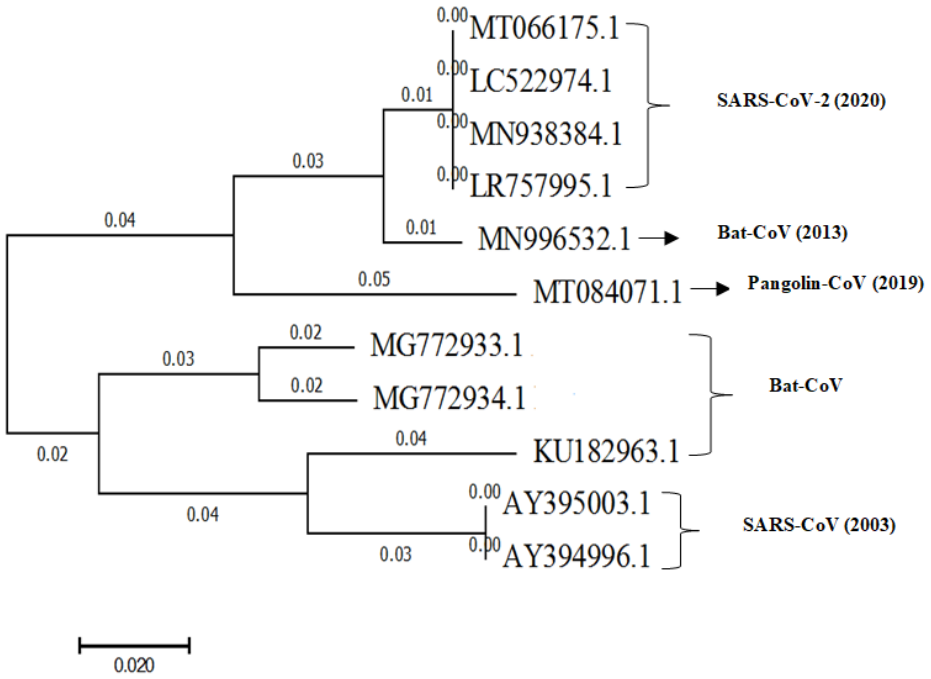

Figure 2. Phylogenetic tree of SARS CoV-2 sequences, Bat coronavirus sequences, Pangolin-CoV and SARS Cov. SARS, severe acute respiratory syndrome; CoV, Corona virus

(AY274119.3) suggest that in 2003 the severe acute respiratory syndrome-related coronavirus were found in Canada Toronto that is found to having close similarity of 99.6% with SARS coronavirus. Accession Number (DQ182595.1, AY8648051) that occur in China in 2004. In 2012 Bat SARS-like coronavirus occur in china accession number (KC881006.1, KY417152.1, KY417146.1 and KY417144.1) have similarity of 95%. Middle East respiratory syndrome-related coronavirus (MG923471.1) in Burkina have 66% identity, occur in 2015. Analysis of Genbank information regarding Middle East respiratory syndrome-related coronavirus (gene bank accession number KJ813439.1) found in 2014 in Saudi Arabia have 66% identity. Middle East respiratory syndrome Batcoronavirus accession number (MG596802.1) genome have 67% identity found in Italy in 2011. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Gene bank accession number (MT007544.1) occur in Australia in 2020 found to have 80% identity. Complete genome information of severe acute respiratory syndrome-related coronavirus (DQ898174) in Canada year 2009 show 99% identity. SARS coronavirus (gene bank accession number AY323977) in Italy in 2003 have 99% identity.

In the analysis we found also distantly related MERS genomes from Gene Bank entries (KJ813439.1, MG923470.1 and MG923471.1). All the humans with SARS-CoV-2 appeared very identical to each other, despite of sampling from different localities. Bat coronavirus appears to be the closest species.

Event matrix of vicariance and dispersal were obtained using Bayesian Binary MCMC (BBM) analysis with following probability algorithms. BBM analysis suggests that 14 global dispersal were noted with vicariance of 7 and no extension were found to occur. At node 18, 19 one events of vicariance and two events of dispersal occur at both node with a probability of .6331 and 0.5026 respectively (Figure 1). One events of vicariance and two dispersal events were noted at node 21, node 23, node 26 and node 27 with probability of 0.70, 0.54, 0.069 and 0.20 respectively. There is no vicariance and dispersal event at node 15, 16, 17, 20 and 24 with probability rate of 0.99 (Figure 1). Dispersal to the Canada (C) and Italy (B) took place (nodes 19 and 20). The tree data in Figure 1 suggests that dispersal occur at node 18 from China and Italy. Presence of node 26 show dispersal event to Saudi Arabia (D) and Australia (G) (Figure 1).

The whole genome of corona virus phylogenetic analysis revealed that SARS-CoV-2 is closely related (96% nucleotide similarity) to a group of Bat coronaviruses (accession number MN996532) and have identity of 91% with pangolin corona virus. In phylogenetic trees based on the nucleotide sequences of whole genomes and viral protein genes, the SARS-CoV-2 clustered (82% nucleotide similarity) with the 2002–2003 SARS-CoV pandemic. Significant similarities were noted in nucleotide sequence and predicted structure between SARS-CoV

and SARS-CoV-2 (Figure 2).

In Figure 2 the phylogenetic tree were presented with analysis of different coronaviruses including SARS-CoV, bat-CoV, pangolin-Cov and recently SARS-CoV-2. Four strain of SARS-CoV-2 are on the same clade as bat coronavirus, which possibly suggest the closest relation of SARS-CoV-2 with bat-coronavirus.

Description of the SARS-CoV-2 genome and analyses of the phylogenetic relationships between SARS-CoV-2, bat SARS-related CoVs and SARS-CoV, observation show a close phylogenetic relationship (96% identical at the whole genome level), suggesting a bat origin of SARS-CoV-2[17]. However, additional studies are required to conclude whether there is intermediate host in virus transmission to humans. In summation, these study provide genomic variation and identity among SARS-CoV-2, bat SARS-related CoVs and SARS-CoV.

We retrieve 25 full genomic sequences from the NCBI gene bank which include Gene Bank deposited sequence of 11 corona virus and their identical genome and 14 other full genome sequence were retrieve that include Bat corona virus and corona virus associated with camel. Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV from China (MN938384.1, MT066175.1, and LC522974.1) and SARS Cov-2 from the Wuhan seafood market pneumonia virus isolate (LR757995.1) have found to be closely similarity with other related virus. To compare SARS Cov-2 with closely identical viral species, we obtained four sequences of Bat SARS-like coronavirus genomes from Gene Bank (Accession No: KU182963.1, MG772934.1, MG772933.1, and MN996532.1). 2019-nCoV have significantly transmitted through air flights from country to other neighboring countries and the consequences of the virus are likely to be higher because of the travelling[18].

4 Discussion

Current analysis of our results highlight a high level of identity within SARS CoV-2 genomes and a clear origin from other beta Coronaviruses, especially SARS, MERS and Bat-coronavirus. Our result suggested that the previous results highlighting the Bat-coronavirus as a likely evolutionary link with the SARS viruses and the recent pandemic SARS CoV-2[4], same result were also predicted in previous study[19].

Our result showed that SARS-CoV-2 has about 96% genome identity with the bat coronavirus that show the possibility of a zoonotic origin. Zhou et al present the same result that 2019-nCoV were found to have 96% similarity with genome of bat coronavirus[20]. 2019-nCoV are most closely related to bat-CoV, which suggested that bats might be the original host for 2019-novel coronavirus[21]. However it is early to predict the origin of the 2019-nCoV with lack of comprehensive analysis of 2019-nCoV strains from different geographical location of the world. Bats are thought to be reservoirs of coronaviruses. It is evaluated that bats which are either experimentally infected or naturally infected do not show clinical signs of disease. From these clarifications researchers guess that bats are the likely reservoirs for coronavirus[22]. Lack of forest might be associated with interaction of bats with human. It is also difficult to conduct study on bats due to unidentified all species, lack of reagent, high risk of infection, method and expertise for studying bats[23].

The genetic similarity among SARS CoV-2, SARS, MERS and Bat-coronavirus is very high, with identities of above 85%, with full conservation of the genome length. There is likely the transmission chain began from the bat and reached the human. Same result were also suggested by Benvenuto *et al.*[4] and also declared that, the structural analysis of two important viral proteins, the spike like nucleoprotein and nucleocapsid protein established the significant similarity of the SARS CoV-2 with the bat-like SARS CoV and its variation from SARS coronavirus.

The structural N protein is involved in virion assembly, playing a significant role in transcription and assembly of virus. Mutation in these proteins enhance the infectivity rate but lower the mortality rate of SARS CoV-2 from SARS CoV virus that cause epidemic in 2002. Ji *et al.*[24] clarify homologous recombination within the spike glycoprotein of SARS CoV-2 resulted in cross species transmission and suggested snake as probable reservoir of virus for human infection. In previous study, it has been prove that mutation pressure, compositional properties, gene expression, dinucleotide and natural selection affect the codon usage bias of Bungarus specie[25].

These result, along with our study analysis, supporting bat origin of infection, could explain the transmission dynamics of the SARS CoV-2.

Conflicts of interest

The authors declare that they have no competing interests.

Authors contributions

FUH, KA, MR and SUR analyzed the data and drafted the manuscript, while SU, IA, HM and WU conceived the study, was involved in data analysis, and critically reviewed the manuscript. SS, MI and AG revised the manuscript. All authors approved the manuscript.

References

[1] Cui J, Li F and Shi ZL. Origin and evolution of pathogenic coronaviruses. Nature Reviews Microbiology, 2019, **17**(3):

181-192.
https://doi.org/10.1038/s41579-018-0118-9

[2] Schoeman D and Fielding BC. Coronavirus envelope protein: current knowledge. Virology Journal, 2019, **16**(1): 69.
https://doi.org/10.1186/s12985-019-1182-0

[3] Hu B, Ge X, Wang LF, *et al*. Bat origin of human coronaviruses. Virology Journal. 2015, **12**(1): 221.
https://doi.org/10.1186/s12985-015-0422-1

[4] Benvenuto D, Giovanetti M, Ciccozzi A, *et al*. The 2019-new coronavirus epidemic: evidence for virus evolution. Journal of Medical Virology, 2020, **92**(4): 455-459.
https://doi.org/10.1002/jmv.25688

[5] Lu R, Zhao X, Li J, *et al*. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet, 2020, **395**(10224): 565-574.
https://doi.org/10.1016/S0140-6736(20)30251-8

[6] Haq FU, Roman M, Ahmad K, *et al*. Artemisia annua: trials are needed for COVID-19. Phytotherapy Research. 2020.

[7] Khan MU, Muazzam A, Farooq H, *et al*. Interferon-β: treatment option against COVID-19. Geriatric Care, 2020, **6**(3): 65-66.
https://doi.org/10.4081/gc.2020.9093

[8] Aljofan M and Gaipov A. COVID-19 Treatment: The Race Against Time. Electronic Journal of General Medicine, 2020, **17**(6): em227.
https://doi.org/10.29333/ejgm/7890

[9] Jeanmougin F, Thompson JD, Gouy M, *et al*. Multiple sequence alignment with CLUSTAL X. Trends in Biochemical Sciences, 1998, **23**(10): 403-405.
https://doi.org/10.1016/S0968-0004(98)01285-7

[10] Altschul SF, Gish W, Miller W, *et al*. Basic local alignment search tool. Journal of Molecular Biology, 1990, **215**(3): 403-410.
https://doi.org/10.1016/S0022-2836(05)80360-2

[11] Kumar S, Stecher G, Li M, *et al*. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution, 2018, **35**(6): 1547-1549.
https://doi.org/10.1093/molbev/msy096

[12] Tamura K and Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution, 1993, **10**(3): 512-526.
https://doi.org/10.1093/oxfordjournals.molbev.a040023

[13] Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985, **39**(4): 783-791.
https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

[14] Ronquist F and Hülsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 2003, **19**: 1572-1574.
https://doi.org/10.1093/bioinformatics/btg180

[15] Yan Y, Harris AJ and Xingjin H. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. Molecular Phylogenetics and Evolution, 2010, **56**: 848-850.
https://doi.org/10.1016/j.ympev.2010.04.011

[16] Yu Y, Harris AJ, Blair C, *et al*. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Molecular phylogenetics and evolution, 2015, **87**: 46-49.
https://doi.org/10.1016/j.ympev.2015.03.008

[17] York A. Novel coronavirus takes flight from bats? Nature Reviews Microbiology, 2020, **18**: 191.
https://doi.org/10.1038/s41579-020-0336-9

[18] Haider N, Yavlinsky A, Simons D, *et al*. Passengers' destinations from China: low risk of Novel Coronavirus (2019-nCoV) transmission into Africa and South America. Epidemiology & Infection, 2020, **148**(41): 1–7.
https://doi.org/10.1017/S0950268820000424

[19] Ceraolo C and Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. Journal of Medical Virology. 2020, **92**(5): 522-528.
https://doi.org/10.1002/jmv.25700

[20] Zhou P, Yang XL, Wang XG, *et al*. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. BioRxiv, 2020.
https://doi.org/10.1101/2020.01.22.914952

[21] Lu R, Zhao X, Li J, *et al*. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet, 2020, **395**(10224): 565-574.
https://doi.org/10.1016/S0140-6736(20)30251-8

[22] Banerjee A, Kulcsar K, Misra V, *et al*. Bats and coronaviruses. Viruses. 2019, **11**(1): 41.
https://doi.org/10.3390/v11010041

[23] Schountz T. Immunology of bats and their viruses: challenges and opportunities. Viruses, 2014, **6**(12): 4880-4901.
https://doi.org/10.3390/v6124880

[24] Ji W, Wang W, Zhao X, *et al*. Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. Journal of Medical Virology, 2020, **92**(4): 433-440.
https://doi.org/10.1002/jmv.25682

[25] Chakraborty S, Nag D, Mazumder TH, *et al*. Codon usage pattern and prediction of gene expression level in Bungarus species. Gene, 2016, **604**: 48-60.
https://doi.org/10.1016/j.gene.2016.11.023