Deep phylogenetic analysis of *Orthocoronavirinae* genomes traces the origin, evolution and transmission route of 2019 novel coronavirus

Amresh Kumar Sharma and Anup Som*

Centre of Bioinformatics
Institute of Interdisciplinary Studies
University of Allahabad
Prayagraj – 211002, India

*Corresponding Author:
Email: som.anup@gmail.com
Website: www.somlab.in

Abstract

The outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Wuhan city, China in December 2019 and thereafter its spillover across the world has created a global pandemic and public health crisis. Today, it appears as a threat to human civilization. Scientists and medical practitioners across the world are involved to trace out the origin and evolution of SARS-CoV-2 (also called 2019 novel coronavirus and referred as 2019-nCoV), its transmission route, cause of pathogenicity, and possible remedial action. In this work, we aim to find out the origin, evolutionary pattern that led to its pathogenicity and possible transmission pathway of 2019-nCoV. To achieve the aims we conducted a large-scale deep phylogenetic analysis on the 162 complete *Orthocoronavirinae* genomes consisting of four genera namely *Alphacoronavirus*, *Betacoronavirus*, *Deltacoronavirus* and *Gammacoronavirus*, their gene trees analysis and subsequently genome and gene recombination analyses. Our analyses revealed that i) bat, pangolin and anteater are the natural hosts of 2019-nCoV, ii) outbreak of 2019-nCoV took place via inter-intra species transmission mode, iii) host-specific adaptive mutation made 2019-nCoV more virulent, and iv) the presence of widespread recombination events led to the evolution of new 2019-nCoV strain and/or could be determinant of its pathogenicity.

Keywords: Orthocoronavirinae; 2019-nCoV; Genome/Gene phylogeny; Adaptive mutation; Recombination; Transmission pathway, COVID-19

Highlights

- *Orthocoronavirinae* genome phylogeny revealed that bat, pangolin and anteater are natural reservoir hosts of novel coronavirus (2019-nCoV/SARS-CoV-2).
- Host-specific adaptive mutation occurred among the coronaviruses.
- Transmission of 2019-nCoV to human took place by inter-intra species mode of transmission.
- Presence of widespread recombination events led to the evolution of new 2019-nCoV strain and/or could be determinant of its pathogenicity.

Introduction

Coronaviruses are single-stranded RNA virus of 26 to 32 kilobases (kb) nucleotide chain and consists of both structural and non-structural proteins. They have been known to cause lower and upper respiratory diseases, central nervous system infection and gastroenteritis in a number of avian and mammalian hosts, including humans (Gorbalenya et al 2020). The recent outbreak of novel coronavirus (2019-nCoV/SARS-COV-2) associated with acute respiratory disease called coronavirus disease 19 (commonly known as COVID-19) has caused a global pandemic and has spread over 212 countries (WHO, COVID-19 situation report). As of now, more than 4 million people have been infected and approximately 300 thousand people have died. Today, COVID-19 appears as a global threat to public health as well as to the human civilization (WHO, COVID-19 situation reports). As it was initially outbreak in Wuhan city, Hubei province, China in December 2019 but then rapidly spread to several European countries and subsequently almost the entire world (Wu et al., 2020; Zhu et al., 2019).

Coronaviruses are placed within the family *Coronaviridae*, which has two subfamilies namely *Orthocoronavirinae* and *Torovirinae*. *Orthocoronavirinae* has four genera: *Alphacoronavirus* (average genome size 28kb), *Betacoronavirus* (average genome size 30kb), *Gammacoronavirus* (average genome size 28kb), and *Deltacoronavirus* (average genome size 26kb) (de Groot et al. 2011). Coronaviruses are typically harbored in mammals and birds. Particularly *Alphacoronavirus* and *Betacoronavirus* infect mammals, and *Gammacoronavirus* and *Deltacoronavirus* infect avian species (Woo et al., 2009; 2010; Fan et al., 2019).

The previous important outbreaks of coronaviruses are severe acute respiratory syndrome coronavirus (SARS-CoV) outbreak in China in 2002/03, Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in 2012 that resulted severe epidemics in the respective geographical regions (Eickmann et al., 2003; Vijaykrishna et al., 2007; Zumla et al, 2015; Hayes et al., 2019). The present outbreak of 2019-nCoV is the third documented spillover of an animal coronavirus to humans in only two decades that has resulted in a major pandemic (Velavan and Meyer, 2020; Lai et al., 2020). Despite the deadly infection caused by 2019-nCoV, till are no specific vaccines or medicines for COVID-19.

Scientific communities across the world are trying to understand several fundamental and applied questions such as: What is the origin of 2019-nCoV? What are the possible transmission routes? Why 2019-nCoV is more deadly than other CoVs? What is its possible clinical diagnosis & treatment? etc. Consequently, a large number of research outcomes are being consistently published. In this paper, we aim to find out the origin and evolution of 2019-nCoV, and its possible transmission pathway through deep phylogenetic analysis.

Materials and Methods

Data selection

162 *Orthocoronavirinae* genomes were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/) and Virus Pathogen Database and Analysis Resource (https://www.viprbrc.org/). We only considered complete genome sequences having no unidentified nucleotide characters. Our dataset included 23 *Alphacoronavirus*, 92 *Betacoronavirus*, 32 *Deltacoronavirus* and 15 *Gammacoronavirus* genomes from different subgenera, diverse host species and from wide geographical location (Cui et al., 2019). Further for rooting the tree, we used two genome sequences from *Torovirus* and two from *Bafinivirus* belonging to domestic cow and fish respectively. Three random sequences were generated by using the average genome length of the *Orthocoronavirinae* genomes and their average GC%. The random sequences were used to check the reliability of the topology. Overall, the phylogenetic analysis consists of 162 complete *Orthocoronavirinae* genomes; four outgroups sequences and three random sequences.

Phylogenetic reconstruction

The genome sequences were aligned using the MAFFT alignment tool (Katoh et al., 2002). Genome tree of the *Orthocoronavirinae* and *Betacoronavirus* were reconstructed using maximum likelihood (ML) method and GTR+G+I model of sequence evolution as revealed by the model test with 1000 bootstrap support. Trees were reconstructed using IQ-TREE software (Nguyen et al., 2015) and were visualized with iTOL software (Letunic et al., 2019). Five gene trees namely Orf1ab, Spike (S), Membrane (M), Envelope (E) and Nucleocapsid (N) were reconstructed using amino acid sequences. ML method of tree reconstruction and protein-specific amino acids

evolution model as revealed by the model test was used for gene trees reconstruction. Bootstrap test with 1000 bootstrap replicates was carried out to check the reliability of the gene trees.

Genome and gene recombination analysis

Potential recombination events in the history of the *Betacoronaviros* were assessed using the RDP5 package (Martin et al., 2015). The RDP5 analysis was conducted based on the complete genome sequence using RDP, GENECONV, BootScan, MaxChi, Chimera, SiScan, and 3Scan methods. Putative recombination events were identified with a Bonferroni corrected P-value cut-off of 0.05 supported by more than four methods.

Results and Discussion

The genome phylogeny of *Orthocoronavirinae* depicts that Alpha, Beta, Delta and Gamma coronaviruses clustered according to their cladistic relations (Fig. 1). This result is consistent with the other results (Luk et al. 2019; Wu et al., 2020). Furthermore, *Gammacoronavirus* and *Deltacoronavirus* appeared in a single clade (i.e. formed a monophyletic group). Interestingly *Betacoronavirus* appeared as paraphyletic clade to the Gamma-Delta monophyletic clade. *Alphacoronavirus* emerged as a basal radiation of the *Orthocoronavirinae* phylogeny. Further, deep analysis of the genome tree revealed that the same host strains from different geographical location of *Alphacoronaviruses* are conserved (Li et al., 2020). This is probably due to the host specific adaptive mutations (Songa et al., 2005; Andersen et al., 2020). For example,

Figure 1: *Orthocoronavirinae* genome phylogeny. The genome tree consists of 162 complete *Orthocoronavirinae* genomes, four outgroups and three random sequences. Alignment consists of 65,157bp aligned nucleotide characters (5,769bp completely aligned characters). Tree was reconstructed using ML method with GTR+G+I model of nucleotide evolution along with 1000 bootstrap replicates (asterisk at the nodes represent support values >70%). Tree was rooted with the four *Torovirinae* genomes. Most recent common ancestors (MRCA) are the closely related ancestor group from 2019-nCoV.

Alphacoronavirus strains from ferret Japan and ferret Netherland is monophyletic. Similarly cat UK is monophyletic to cat Netherland, and human China is monophyletic to human Netherland. Another interesting outcome is all *Alphacoronavirus* camel strains of Saudi Arabia appeared in a distinct sub-clade where bat Ghana appeared as outgroup which clearly indicates interspecies transmission took place from bat (Ghana) to camel followed by adaptive mutation (York, 2020; Zhou et al., 2020).

Deltacoronavirus and *Gammacoronavirus* clade clearly exhibits a similar evolutionary pattern. In case of *Deltacoronaviruses*, swine Vietnam and swine Hong Kong share a single common ancestor. Similarly, swine China and swine South Korea are monophyletic clade and swine Japan is monophyletic to swine South Korea. In case of *Gammacoronaviruses* (whose natural hosts are avian species), chicken Peru and chicken Uruguay shared a single common ancestor. Similarly, chicken Iraq is monophyletic to chicken Egypt strain. These results clearly confirm that coronavirus strains are present in a large number of hosts those are widespread in different geographical location.

In this work, we considered 92 complete *Betacoronavirus* genomes belonging to five subgenera namely Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus and Sarbecovirus (2019-nCoV/SARS-COV-2 belongs to this group). Phylogenetic analysis of *Betacoronavirus* genomes revealed that the five subgenera clustered separately (Fig. 2). Further, the *Betacoronavirus* genome tree depicts that the host-specific strains from distance geographical locations formed monophyletic clades. For example, in Embecovirus clade, strain BJ01 P9 human China is monophyletic to the Caen1 human France strain. Similarly, Embecovirus B1 24F buffalo Bangladesh is monophyletic to BCV AKS 01 cattle China. This result infers that irrespective of the hosts' geographical location/diversity host-specific adaptive mutation occurred.

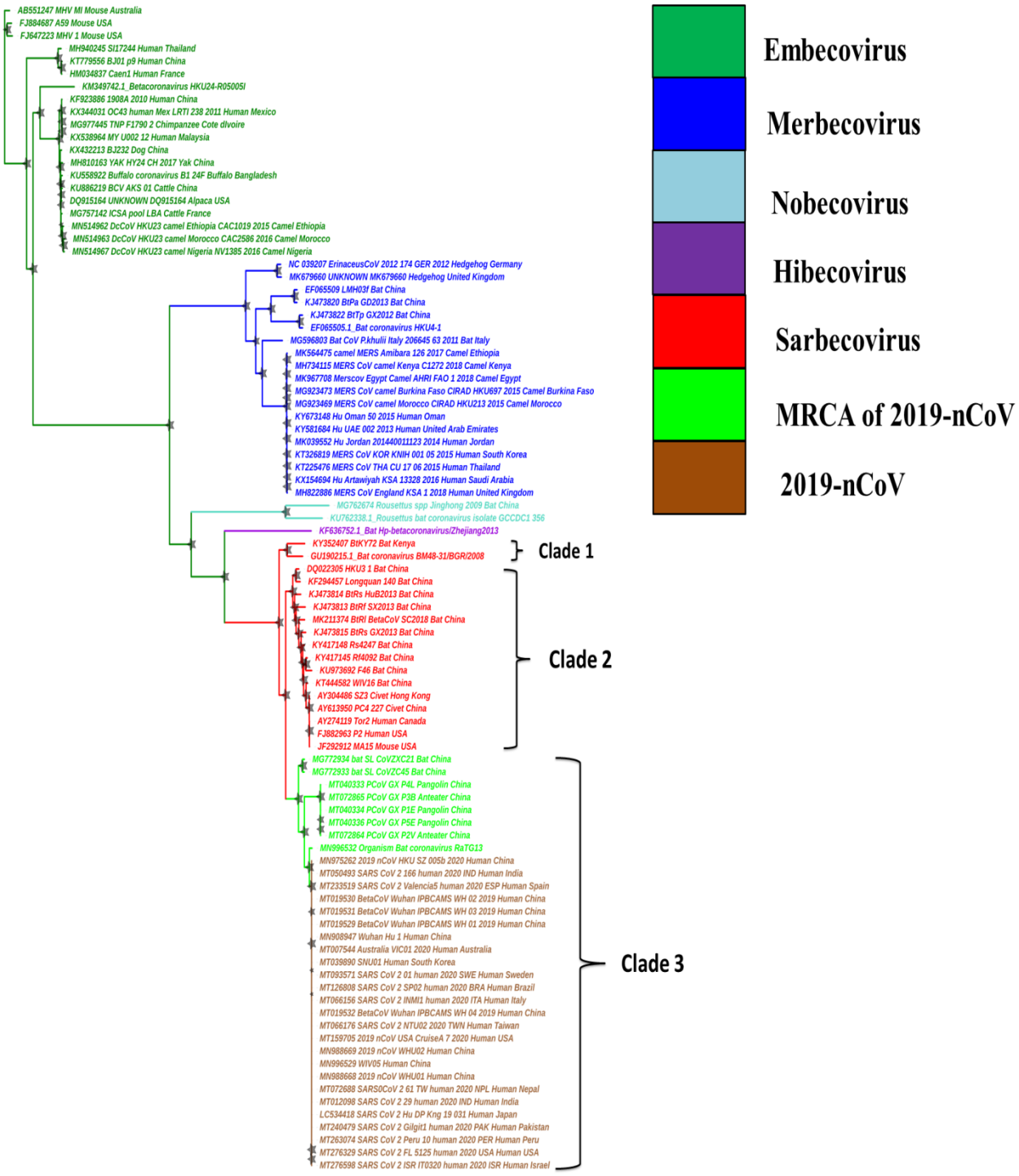

Figure 2: *Betacoronavirus* genome phylogeny. The genome tree consists of 92 complete *Betacoronavirus* genomes. Alignment consists of 41,054bp aligned nucleotide characters (23,064bp completely aligned characters). Tree was reconstructed using ML method with GTR (GTR+G+I) model of nucleotide evolution along with 1000 bootstrap replicates (asterisk indicates value >70%). Most recent common ancestors (MRCA) are the closely related ancestor group from 2019-nCoV.

Novel coronavirus (2019-nCoV) belongs to Sarbecovirus subgenus (Fig. 2). Sarbecoviruses formed three distinct clades, where Clade 1 consists of only bat as host species. In Clade 2, host species are bat, human and civet. Similarly, in Clade 3 the host species are bat, human, pangolin and anteater, and it depicts bat-CoV-RaTG13 (NCBI Acc no. Mn996532) is closest to CoV humans, which was evolved by the recombination of bat-SL-CoVZXC21 and bat-SL-CoVZC45 strains. The clades analysis clearly asserts bat, pangolin and anteater are the natural reservoir of 2019-nCoV, and transmission from bat /pangolin/anteater to humans took place through intermediate organisms (Cui et al., 2019; York, 2020). Furthermore, this phylogenetic tree reveals *Betacoronavirus* sequences are conserved in their respective hosts after acquiring adaptive mutation (e.g. all bat hosts clustered in Clade 2 and human hosts are in Clade 3) (Lu et al., 2020).

In addition to genome phylogeny, gene tree analysis was also conducted as it provides a more reliable basis for studying genome evolution. Five gene trees namely orf1ab, spike (S), envelope (E), Membrane (M) and Nucleocapsid (N) were used for gene tree analysis (Figs. 3-7). Except N gene tree (Fig. 6), other four gene trees have shown similar evolutionary pattern with respect to their subgenera and were in concordance with their genome tree. This observation implies that the *Betacoronavirus* genome evolution is influenced by the genes' evolution. Further deep analysis found, though subgenera-wise four gene trees are similar, but within subgenera there are widespread phylogenetic incongruences (Jeffroy et al., 2006). This result led us to hypothesize that HGTs and/or recombination had occurred among *Betacoronaviruses* in the past that are caused to evolve new strains including the emergence of pathogenic lineages. However, it was found that N gene tree has significant topological difference with other gene trees. This might be possible as gene tree differs species tree for various analytical and/or biological reasons (Degman et al., 2009; Som, 2013; 2015).

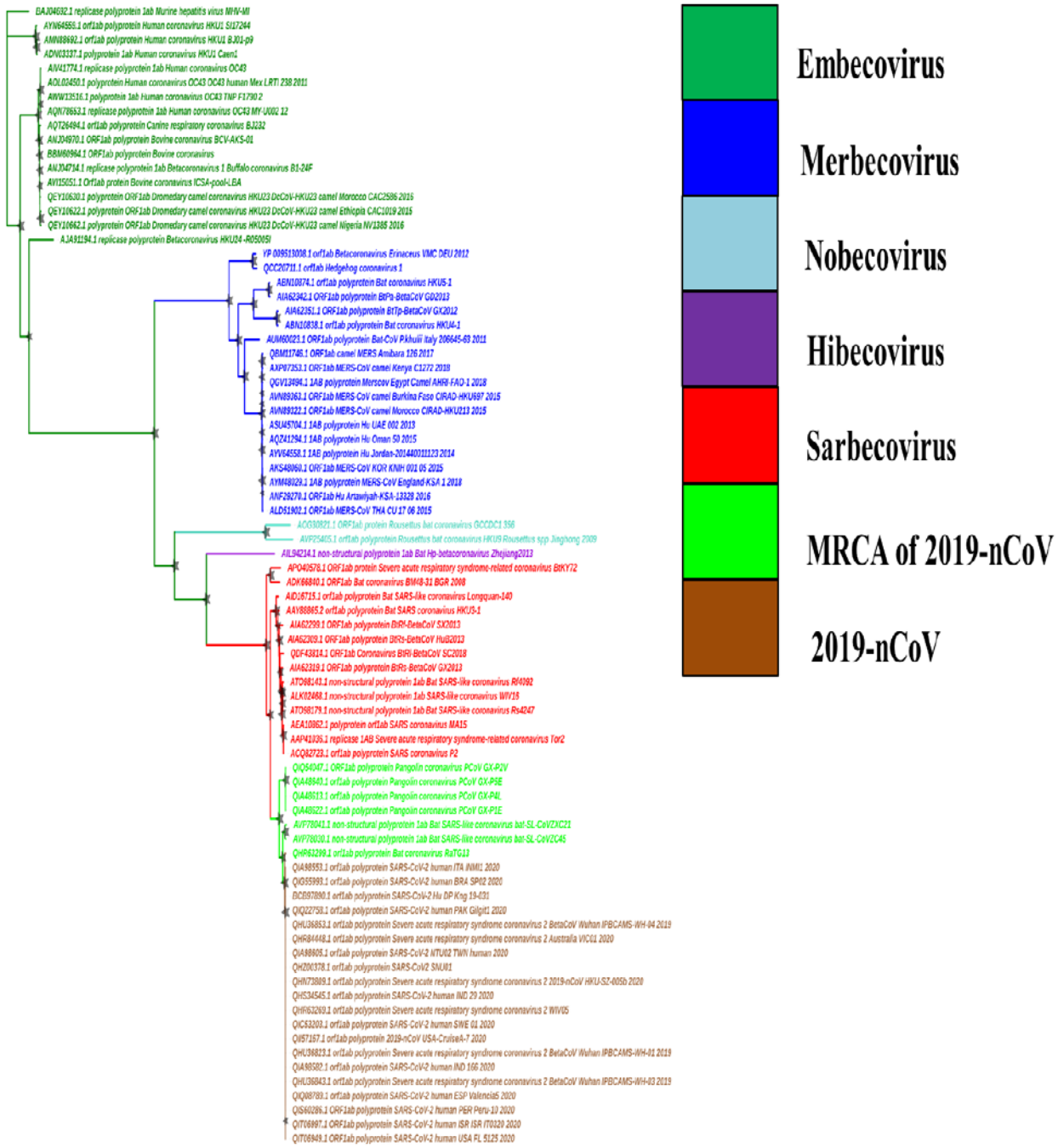

Figure 3: Orf1ab gene phylogeny. Alignment consists of 8,152bp aligned amino acid characters (6,276bp are completely aligned characters). Tree was reconstructed using ML method and LG+I+G4 model of protein evolution along with 1000 bootstrap replicates (asterisk indicates value >70%). MRCA (most recent common ancestors) are the closely related ancestor group from 2019-nCoV.

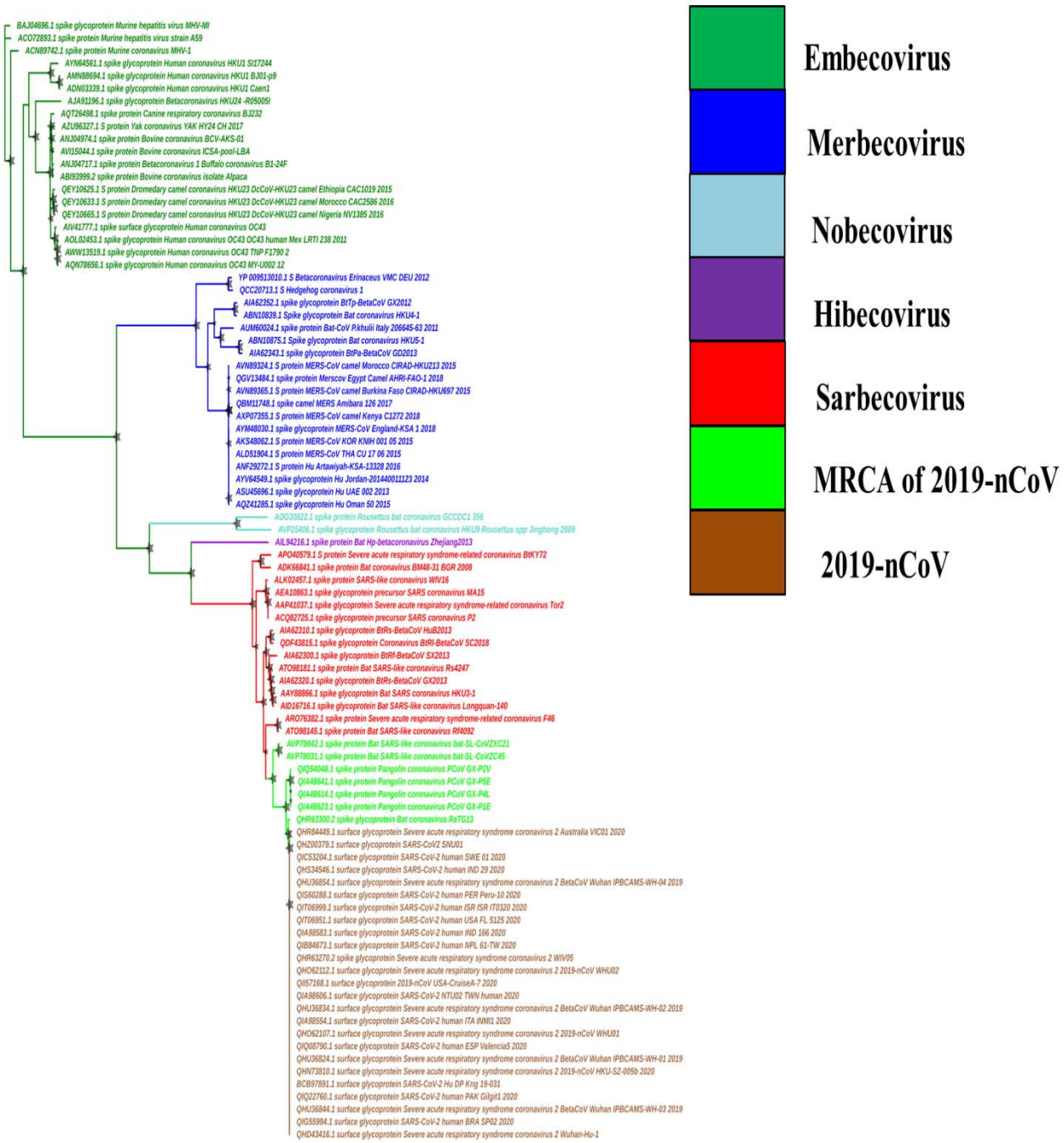

Figure 4: Spike (S) gene phylogeny. Alignment consists of 1,621 aligned amino acid characters (1,071bp are completely characters). Tree was reconstructed using ML method and WAG+I+G4 model of protein evolution along with 100 bootstrap replicates (asterisk indicates value >70%). MRCA (most recent common ancestors) are the closely related ancestor group from 2019-nCoV.

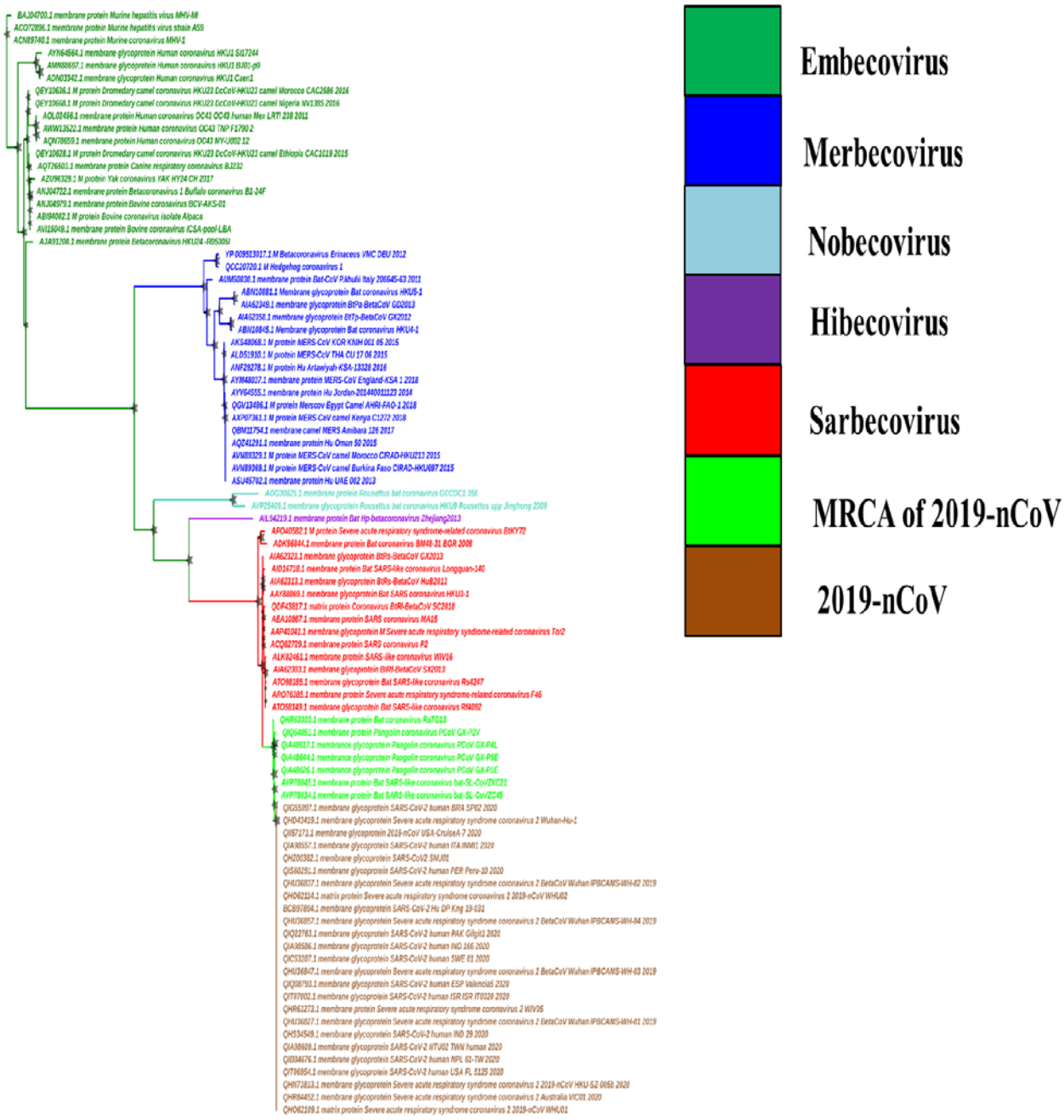

Figure 5: Membrane (M) gene phylogeny. Alignment consists of 233bp aligned amino acid characters (213bp are completely aligned characters). Tree was reconstructed using ML method by and LG+G4 model of protein evolution along with 1000 bootstrap replicates (asterisk indicates value >70%). MRCA (most recent common ancestors) are the closely related ancestor group from 2019-nCoV.

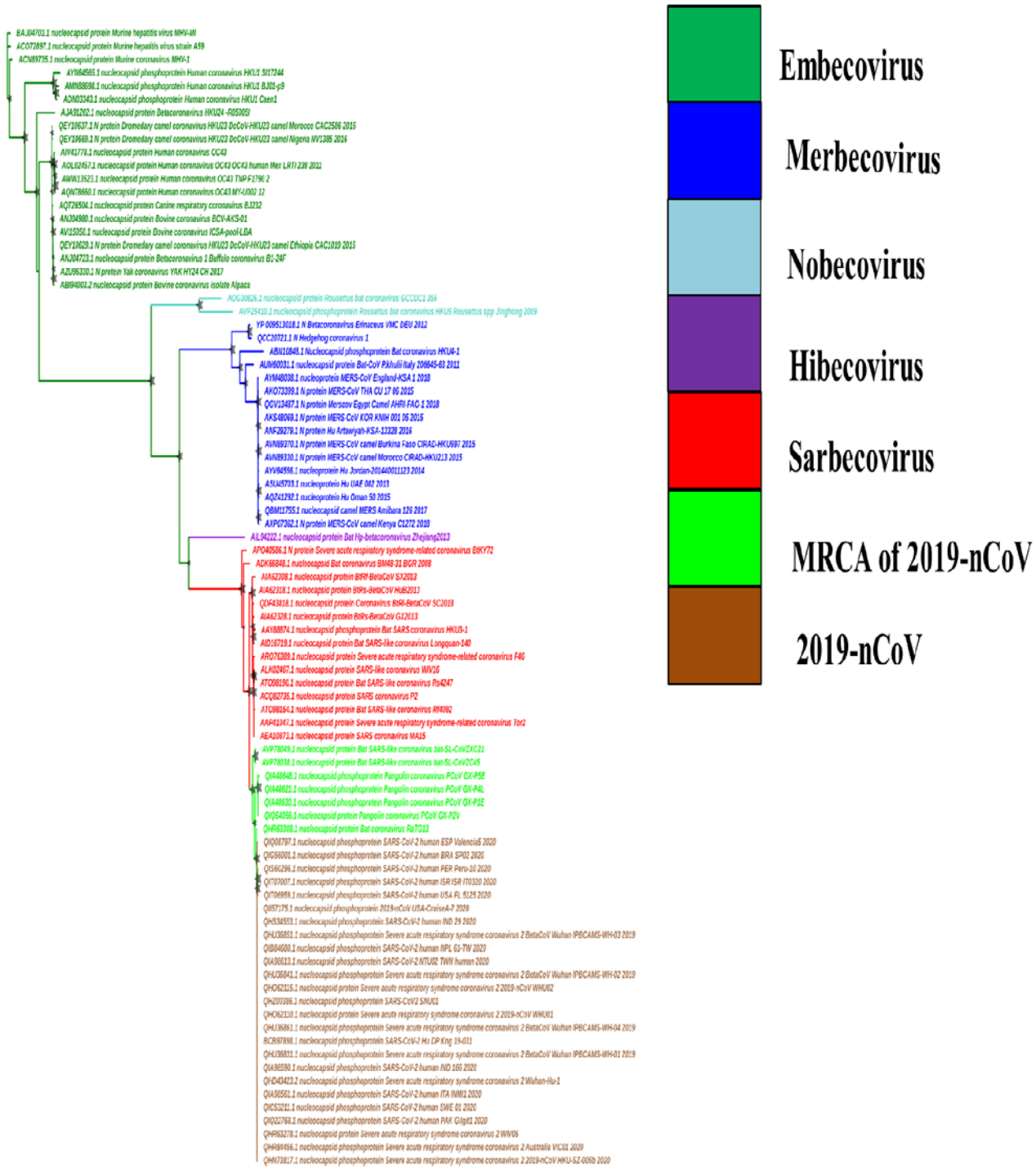

Figure 6: Nucleocaspid (N) gene phylogeny. Alignment consists of 547bp aligned amino acid characters (343bp are completely aligned characters). Tree was reconstructed using ML method and LG+I+G4 model of protein evolution along with 1000 bootstrap replicates (asterisk indicates value >70%). MRCA (most recent common ancestors) are the closely related ancestor group from 2019-nCoV.

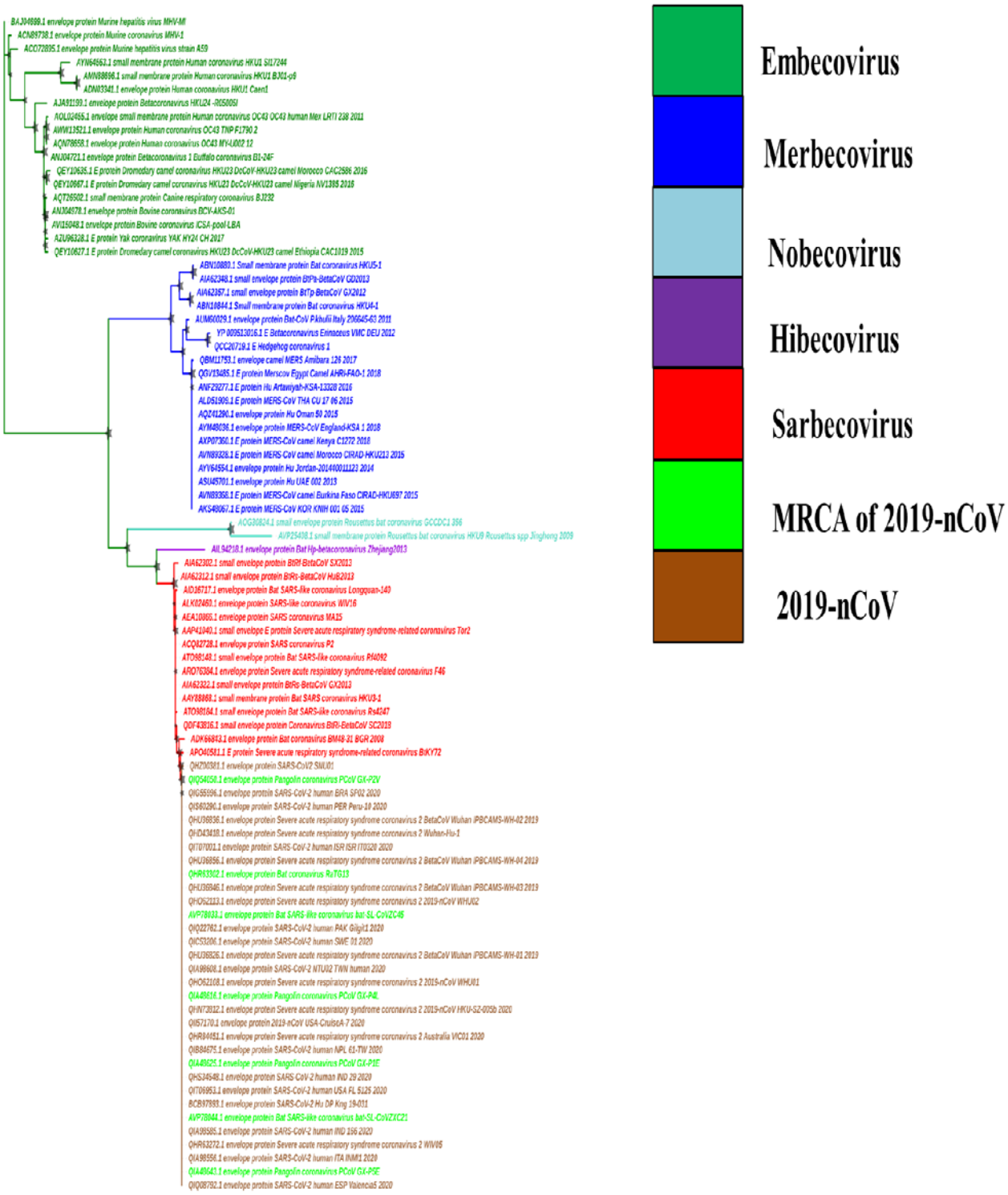

Figure 7: Envelope (E) gene phylogeny. Alignment consists of 90 aligned amino acid characters (74 are completely aligned characters). Tree was reconstructed using ML method and JTT+I+G4 model of protein evolution along with 1000 bootstrap replicates (asterisk indicates value >70%). MRCA (most recent common ancestors) are the closely related ancestor group from nCoV19.

We conducted both genome (*Betacoronavirus*) and gene recombination analysis using RDP5 package (Martin et al., 2015). The genome recombination analysis detected 21 putative recombination signals (Table 1). Recombination results show that major recombination events took place between Clade 2 and Clade 3 of Sarbecovirus (Table 1). Gene recombination analysis found there are widespread recombination events in orf1ab and S proteins (Table 2). M and N proteins reported few recombination events and envelope (E) protein did not show any recombination event. Major genetic variations in these genes, particularly in spike (S) gene, seemed essential for the transition from animal-to-human transmission to human-to-human transmission, which eventually caused the outbreak of 2019-nCoV (Su et al., 2016; Luk et al. 2019; Jaimes et al., 2020). Details of the genome and gene recombination analyses are given in Table 1 and Table 2 respectively.

It was found that the most closely related bat and SARS-CoVs diverged in 1986, an estimated divergence time of 17 years prior to the outbreak in December 2002 (Vijaykrishna et al., 2007). In a similar line, a recent article by Lu et al. (2020) reported that intermediate transmission happened long back of 2019-nCoV outbreak in December 2019. Incidentally 2019-nCoV (SARS-CoV-2) outbreak also happened in the same time interval of 17 years in 2019 and also in the same winter season. Further, evolutionary studies have shown that 2019-nCoV is genetically distance from SARS-CoVs with 88~96% sequence similarity between them (Ceraolo and Giorgi, 2020; Lu et al., 2020). So the open question is which factors made 2019-nCoV is highly pathogenic: mutation, recombination, structural changes of spike protein, combination of multiple factors, or something else which needs to be explored. Thus, further studies on detailed analysis of 2019-nCoV genomes, particularly the spike protein sequences and structures, as well as the receptors for the individual novel coronaviruses will enable to understand the origin, pathogenicity and mechanism behind interspecies jumping at the molecular level, which will help in the prevention of future zoonotic events.

Table1: Detected recombination events in the Betacoronavirus genomes. Twenty one potential recombination signals were detected by using RDP5 package. A recombination event was reported, when five out of seven methods detected it. Details of genome recombination analysis are given in the text.

Events no.	Recombinant sequences	Major parent	Minor parent
1	MG772933|bat_SL_CoVZC45|Bat|China	MN996532|Organism:Bat coronavirus RaTG13	KF294457|Longquan_140|Bat|China
2	KY417145|Rf4092|Bat|China	KT444582|WIV16|Bat|China	KU973692|F46|Bat|China
3	KF294457|Longquan_140|Bat|China	DQ022305|HKU3_1|Bat|China	MG772934|bat_SL_CoVZXC21|Bat|China
4	KJ473815|BtRs_GX2013|Bat|China	KF294457|Longquan_140|Bat|China	KT444582|WIV16|Bat|China
5	MK211374|BtRl_BetaCoV/SC2018|Bat|China	KF294457|Longquan_140|Bat|China	AY304486|SZ3|Civet|Hong_Kong
6	KY417148|Rs4247|Bat|China	DQ022305|HKU3_1|Bat|China	KY417145|Rf4092|Bat|China
7	KJ473815|BtRs_GX2013|Bat|China	DQ022305|HKU3_1|Bat|China	AY304486|SZ3|Civet|Hong_Kong
8	KU973692|F46|Bat|China	FJ882963|P2|Human|USA	KJ473814|BtRs_HuB2013|Bat|China
9	MK211374|BtRl_BetaCoV/SC2018|Bat|China	KJ473814|BtRs_HuB2013|Bat|China	KY417145|Rf4092|Bat|China
10	MK211374|BtRl_BetaCoV/SC2018|Bat|China	KJ473814|BtRs_HuB2013|Bat|China	FJ882963|P2|Human|USA
11	MT039890|SNU01|Human|South_Korea	MG772934|bat_SL_CoVZXC21|Bat|China	JF292912|MA15|Mouse|USA
12	MN514962|DcCoV_HKU23/camel/Ethiopia/CAC1019/2015|	MN514963|DcCoV_HKU23/camel/Morocco/CAC2586/2016	KX432213|BJ232|Dog|China
13	KJ473815|BtRs_GX2013|Bat|China	KY417145|Rf4092|Bat|China	DQ022305|HKU3_1|Bat|China
14	FJ647223|MHV_1|Mouse|USA	FJ884687|A59|Mouse|USA	AB551247|MHV_MI|Mouse|Australia
15	KJ473813|BtRf_SX2013|Bat|China	KJ473814|BtRs_HuB2013|Bat|China	JF292912|MA15|Mouse|USA
16	MK211374|BtRl_BetaCoV/SC2018|Bat|China	KY352407|BtKY72|Bat|Kenya	MN996532|Organism:Bat coronavirus RaTG13
17	MN996532|Organism:Bat coronavirus RaTG13	MT040334|PCoV_GX_P1E|Pangolin|China	KY417145|Rf4092|Bat|China
18	KF294457|Longquan_140|Bat|China	KJ473814|BtRs_HuB2013|Bat|China	MG772934|bat_SL_CoVZXC21|Bat|China
19	AY304486|SZ3|Civet|Hong_Kong	KY417148|Rs4247|Bat|China	KU973692|F46|Bat|China
20	KY417148|Rs4247|Bat|China	AY304486|SZ3|Civet|Hong_Kong	KJ473815|BtRs_GX2013|Bat|China
21	KJ473815|BtRs_GX2013|Bat|China	AY613950|PC4_227|Civet|China	MN996532|Organism:Bat coronavirus RaTG13

Table2: Detected recombination events in the four Betacoronavirus genes. A recombination event was reported, when five out of seven methods detected by RDP5 package. Details of gene recombination analysis are given in the text.

Gene name	Recombinant sequence(s)	Major parent sequence	Minor parent sequence
Orf1ab gene	AID16715.1_[Bat_SL-Cov_Longquan140]	AIA62309.1_[BtRs-BetaCoV_HuB2013]	AVP78041.1_[Bat_SL-CoVZXC21
	AVP78030.1_[Bat_SL-CoVZC45]	QHR63299.1_[Bat_coronavirus_RaTG13]	AIA62319.1_[BtRs-BetaCoV_GX2013]
	AVP78041.1_[Bat_SL-CoVZXC21		
	AID16715.1_[Bat_SL-Cov_Longquan140]	AVP78041.1_[Bat_SL-CoVZXC21	AAY88865.2_[Bat_SARS_CoV_HKU3-1]
	QDF43814.1_[Coronavirus_BtRl-BetaCoV_SC2018]	AIA62309.1_[BtRs-BetaCoV_HuB2013]	ALK02468.1_[SARS-like_CoV_WIV16]
Spike (S) gene	ARO76382.1_[SARS_CoV_F46	Unknown (AIA62300.1_[BtRf-BetaCoV_SX2013])	ATO98181.1_[Bat_SL_CoV_Rs4247]
	QDF43815.1_[BtRl-Beta_CoV_SC2018]		
	AVP78042.1_[Bat_SL_CoV_ZXC21	QHR63300.2_[Bat_coronavirus_RaTG13]	AIA62300.1_[BtRf-BetaCoV_SX2013]
	AAY88866.1_[Bat_SARS_CoV_HKU3-1]	ATO98181.1_[Bat_SL_CoV_Rs4247]	Unknown (AAP41037.1_[SARS_CoV_Tor2])
	AIA62300.1_[BtRf-BetaCoV_SX2013]	AID16716.1_Bat_SL_CoV_Longquan140]	Unknown (AAP41037.1_[SARS_CoV_Tor2])
	ALK02457.1_[Bat_SL_CoV_WIV16]	AIA62320.1_[BtRs-BetaCoV_GX2013]	AQZ41285.1_[Hu_Oman_50_2015]
	AUM60024.1_[Bat-CoV_P.khulii_Italy_206645-63_2011]	AIA62343.1_[BtPa-BetaCoV_GD2013]	AQZ41285.1_[Hu_Oman_50_2015]
	QIA98583.1_[SARS-CoV-2_human_IND_166]	AIA62300.1_[BtRf-BetaCoV_SX2013]	Unknown (AIL94216.1_[Bat_Hp-betacoronavirus_Zhejiang_2013])
	AAP41037.1_[SARS_CoV_Tor2]	QHR63300.2_[Bat_coronavirus_RaTG13]	AIA62320.1_[BtRs-BetaCoV_GX2013]
	AWW13519.1_[Human_coronavirus_OC43]	AVI15044.1_[Bovine_coronavirus]_ICSA-pool-LBA]	QEY10625.1_[DcCoV-HKU23]
Nucleocaspid (N) gene	ACN89735.1_[Murine_coronavirus_MHV-1]	AJA91202.1_[Betacoronavirus_HKU24_R05005]	Unknown (AVP25410.1_[Rousettus_Bat_CoV_HKU9_Jinghong_2009])
Membrane (M) gene	AVP25409.1_[Rousettus_Bat_CoV_HKU9_Jinghong_2009]	AVI15049.1_[Bovine_coronavirus]_ICSA-pool-LBA]	ADK66844.1_[Bat_coronavirus_BM48-31_BGR_2008]

Concluding remarks

In this work we performed a large-scale genome and gene phylogenetic analyses of the 162 full *Orthocoronavirinae* genomes and their five protein sequences. Our analysis revealed that: i) bat, pangolin and anteater are the natural reservoir host of *Betacoronavirus*, ii) transmission of *Beta coronavirus* to human took place by inter-intra species model (i.e. from bat/pangolin/anteater to intermediate organism and then from intermediate organism to human), iii) host–specific adaptive mutation occurred among the coronavirus strains including 2019-nCoV, iv) gene tree and gene recombination analysis confirmed the widespread presence of recombination events, and v) genome recombination analysis found that recombination events between intra-subgenera are more frequent than inter-subgenera, which possibly led to the evolution of new strains such as 2019-nCov/SARS-CoV-2.

Acknowledgements

This work was partly supported by the Department of Biotechnology (DBT) and University Grants Commission (UGC), India. Thanks to Dr. Rachana Banerjee for several useful discussions. We thank Mr. Arindam Ghosh for editorial supports.

References

Andersen, K.G., et al., 2020. The proximal origin of SARS-CoV-2. Nat Med 26, 450–452. https://doi.org/10.1038/s41591-020-0820-9

Ceraolo, C., Giorgi, F.M., 2020. Genomic variance of the 2019-nCoV coronavirus. J Med Virol 92, 522–528. https://doi.org/10.1002/jmv.25700

Cui, J., Li, F., Shi, Z.-L., 2019. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 17, 181–192. https://doi.org/10.1038/s41579-018-0118-9

De Groot, R.J., Baker, S.C., Baric, R., Enjuanes, L., Gorbalenya, A.E. and Holmes, K.V., 2011. King AMQ, Lefkowitz E, Adams MJ, Carstens EB Family Coronaviridae. Ninth Report of the International Committee on Taxonomy of Viruses, 806-828.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24, 332–340. https://doi.org/10.1016/j.tree.2009.01.009

Eickmann, M., 2003. Phylogeny of the SARS Coronavirus. Science 302, 1504b–11505. https://doi.org/10.1126/science.302.5650.1504b

Fan, Y., Zhao, K., Shi, Z.-L., Zhou, P., 2019. Bat Coronaviruses in China. Viruses 11, 210. https://doi.org/10.3390/v11030210

Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W. and Penzar, D., 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology.5, 536-544. https://doi.org/10.1038/s41564-020-0695-z

Jaimes, J.A., André, N.M., Chappie, J.S., Millet, J.K., Whittaker, G.R., 2020. Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. Journal of Molecular Biology S0022283620302874. https://doi.org/10.1016/j.jmb.2020.04.009

Jeffroy, O., Brinkmann, H., Delsuc, F., et al., 2006. Phylogenomics: the beginning of incongruence? Trends in Genetics 22, 225–231. https://doi.org/10.1016/j.tig.2006.02.003

Katoh, K., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059–3066. https://doi.org/10.1093/nar/gkf436

Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J. and Hsueh, P.R., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. International journal of antimicrobial agents 55, 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res (44), 242–245. https://doi.org/10.1093/nar/gkw290

Li, C., Yang, Y. and Ren, L., 2020. Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species. Infection, Genetics and Evolution, 82, 104285. https://doi.org/10.1016/j.meegid.2020.104285

Lu, R., et al., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet, 395, 565-574.

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. Infection, Genetics and Evolution 71, 21–30. https://doi.org/10.1016/j.meegid.2019.03.001

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution 1. https://doi.org/10.1093/ve/vev003

Nguyen, L.-T., et al., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32, 268–274. https://doi.org/10.1093/molbev/msu300

Som, A., 2013. Genome-scale approach and the accuracy of phylogenetic methods. J Phylogen Evolution Biol 1: 116. https://doi:10.4172/2329-9002.1000116

Som, A., 2015. Causes, consequences and solutions of phylogenetic incongruence. Briefings in Bioinformatics 16, 536–548. https://doi.org/10.1093/bib/bbu015

Song, H.-D et al., 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proceedings of the National Academy of Sciences 102, 2430–2435. https://doi.org/10.1073/pnas.0409608102

Su, S., Wong, G., Shi, W., et al., 2016. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends in Microbiology 24, 490–502. https://doi.org/10.1016/j.tim.2016.03.003

Velavan, T.P., Meyer, C.G., 2020. The COVID-19 epidemic. Trop Med Int Health 25, 278–280. https://doi.org/10.1111/tmi.13383

Vijaykrishna, D., et al., 2007. Evolutionary Insights into the Ecology of Coronaviruses. JVI 81, 4012–4020. https://doi.org/10.1128/JVI.02605-06

Woo, P.C.Y., et al., 2009. Coronavirus Diversity, Phylogeny and Interspecies Jumping. Exp Biol Med (Maywood) 234, 1117–1127. https://doi.org/10.3181/0903-MR-94

Woo, P.C.Y., et al., 2010. Coronavirus Genomics and Bioinformatics Analysis. Viruses 2, 1804–1820. https://doi.org/10.3390/v2081803

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E.C., Zhang, Y.-Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269. https://doi.org/10.1038/s41586-020-2008-3

WHO (World Health Organization), COVID-19 situation reports.
https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

Xu, L., et al., 2016. Detection and characterization of diverse alpha- and betacoronaviruses from bats in China. Virol. Sin. 31, 69–77. https://doi.org/10.1007/s12250-016-3727-3

York, A., 2020. Novel coronavirus takes flight from bats? Nat Rev Microbiol 18, 191–191. https://doi.org/10.1038/s41579-020-0336-9

Zhou, P., et al., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273. https://doi.org/10.1038/s41586-020-2012-7

Zhu, N., Zhang, D., Wang, W., et al., 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 382, 727–733. https://doi.org/10.1056/NEJMoa2001017

Zumla, A., Hui, D.S. and Perlman, S., 2015. Middle East respiratory syndrome. The Lancet, 386(9997), pp.995-1007. https://doi: 10.1016/S0140-6736(15)60454-8

Supplementary data

Supplementary File 1: Details of the 162 *Orthocoronavirinae* genomes and four outgroup sequences used in this study.