

Research



In-silico nucleotide and protein analyses of S-gene region in selected zoonotic coronaviruses reveal conserved domains and evolutionary emergence with trajectory course of viral entry from SARS-CoV-2 genomic data

Adejoke Olukayode Obajuluwa, Pius Abimbola Okiki, Tiwalola Madoc Obajuluwa, Olakunle Bamikole Afolabi

Corresponding author: Adejoke Olukayode Obajuluwa, Biotechnology Unit, Department of Biological Sciences, Afe Babalola University, Ado Ekiti, Ekiti State, Nigeria. ibitayoao@abuad.edu.ng

Received: 29 Jun 2020 - **Accepted:** 20 Jul 2020 - **Published:** 30 Nov 2020

Keywords: SARS-CoV-2, zoonotic CoVs, phylogeny, evolution, conserved domains, viral entry, spike gene

Copyright: Adejoke Olukayode Obajuluwa et al. Pan African Medical Journal (ISSN: 1937-8688). This is an Open Access article distributed under the terms of the Creative Commons Attribution International 4.0 License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cite this article: Adejoke Olukayode Obajuluwa et al. In-silico nucleotide and protein analyses of S-gene region in selected zoonotic coronaviruses reveal conserved domains and evolutionary emergence with trajectory course of viral entry from SARS-CoV-2 genomic data. Pan African Medical Journal. 2020;37(285). 10.11604/pamj.2020.37.285.24663

Available online at: <https://www.panafrican-med-journal.com//content/article/37/285/full>

In-silico nucleotide and protein analyses of S-gene region in selected zoonotic coronaviruses reveal conserved domains and evolutionary emergence with trajectory course of viral entry from SARS-CoV-2 genomic data

Adejoke Olukayode Obajuluwa^{1,&}, Pius Abimbola Okiki¹, Tiwalola Madoc Obajuluwa², Olakunle Bamikole Afolabi³

¹Biotechnology Unit, Department of Biological Sciences, Afe Babalola University, Ado Ekiti, Ekiti

State, Nigeria, ²Department of Media and Communication, Afe Babalola University, Ado Ekiti, Ekiti State, Nigeria, ³Biochemistry Unit, Department of Chemical Sciences, Afe Babalola University, Ado Ekiti, Ekiti State, Nigeria

&Corresponding author

Adejoke Olukayode Obajuluwa, Biotechnology Unit, Department of Biological Sciences, Afe Babalola University, Ado Ekiti, Ekiti State, Nigeria

Abstract

Introduction: the recent zoonotic coronavirus virus outbreak of a novel type (COVID-19) has necessitated the adequate understanding of the evolutionary pathway of zoonotic viruses which adversely affects human populations for therapeutic constructs to combat the pandemic now and in the future. **Methods:** we analyzed conserved domains of the severe acute respiratory coronavirus 2 (SARS-CoV-2) for possible targets of viral entry inhibition in host cells, evolutionary relationship of human coronavirus (229E) and zoonotic coronaviruses with SARS-CoV-2 as well as evolutionary relationship between selected SARS-CoV-2 genomic data. **Results:** conserved domains with antagonistic action on host innate antiviral cellular mechanisms in SARS-CoV-2 include *nsp 11*, *nsp 13* etc. Also, multiple sequence alignments of the spike (S) gene protein of selected candidate zoonotic coronaviruses alongside the S gene protein of the SARS-CoV-2 revealed closest evolutionary relationship (95.6%) with pangolin coronaviruses (S) gene. Clades formed between Wuhan SARS-CoV-2 phylogeny data and five others suggests viral entry trajectory while revealing genomic and protein SARS-CoV-2 data from Philippines as early ancestors. **Conclusion:** phylogeny of SARS-CoV-2 genomic data suggests profiling in diverse populations with and without the outbreak alongside migration history and racial background for mutation tracking and dating of viral subtype divergence which is essential for effective management of present and future zoonotic coronavirus outbreaks.

Introduction

Coronaviruses (CoVs) are enveloped viruses with a positive-sense, single-stranded RNA genome belonging to the coronaviridae family [1]. CoVs are divided into alpha, beta, gamma and delta groups and the beta group is further composed of A, B, C and D subgroups [2]. The virus belongs to the 2B group of the beta-coronavirus family, which includes SARS-CoV and Middle East respiratory

syndrome coronavirus MERS-CoV [3]. Their entry into respiratory and oesophageal routes accounts for mild to severe acute respiratory syndromes which has led to global epidemics with high morbidity, mortality and immense economic losses in affected human populations [4,5]. Encoded within the 3' end of the viral genome are the four main structural proteins of coronavirus particles: spike (S), membrane (M), envelope (E) and nucleocapsid (N) [6] as shown in Figure 1.

Phylogenetic analyses of 15 human CoV whole genomes revealed 2019 novel CoV (2019-nCoV) genome shares highest nucleotide sequence identity with SARS-CoV (79.7%) while its two evolutionarily conserved regions (envelope and nucleocapsid proteins) had sequence homology of 96% and 89.6% with same respectively [3]. Hence, the nomenclature for the novel type of the coronavirus outbreak. Surface proteins which stick out like crown tips (spikes) on coronaviruses binds to host cell receptors-angiotensin converting enzyme 2 (ACE 2) in epithelial cells in hosts. The S1 subunit (N-terminal) of the surface protein facilitates binding to the ACE2 receptor while the S2 subunit (C-terminal) mediates host cell entry through the binding of the viral S protein to human dipeptidyl peptidase 4 (DPP4), marking onset of infection [7,8].

Interestingly, conserved domains of CoVs have been indicated in literatures as vital entry targets in vaccine and drug development [9,10]. However, growing variability and mutational changes in viruses can cause lack of specificity and reduce efficiency of therapeutic measures. Recombination serves central function in virus replication and evolution in viral infections such as HIV, Ebola and MERS [11,12] while molecular mechanisms (RNA fragmentation and trans-esterification reactions) are possible causes of RNA fragments ligation and subsequent increased novel recombination frequency observed among various RNA viruses [13]. Diverse host factors account for a great deal of genome variability in viral recombinants which ranges from multi-resistance to evolutionary novelties [14]. The emergence of

novel viral variants trafficked by humans and animals alike through global travel has remained a constant threat in public health and increasing complexity of host-viral interactivity in viral adaptation and evolution [15].

Methods

Comparison and analyses of conserved domain of 2019-nCoV/SARS-CoV-2 protein: reference number (initial entry with refSeq number NC_045512.1) SARS-CoV-2 was retrieved from National Centre for Biotechnological Information (NCBI) database and query for its conserved domains (CDS) was launched using affiliated resources. Proteins with similar conserved domains were included in the subsequent multiple sequence alignment of spike gene of zoonotic coronaviruses investigated in this study.

Building the spike (S) gene protein candidates from zoonotic coronavirus hosts: the identification of highly related contigs in a set of viral genomes from blast search of SARS-CoV like sequences in NCBI database (included in supplementary data) directed our subsequent search of spike proteins in the selected nine zoonotic coronaviruses. The CoV host spike proteins which were compared with SARS-CoV-2 (QHD43416.1) are: infectious bronchitis virus (IBV) (ADR51590.1), HCoV-229E (AII82124.1), transmissible gastroenteritis virus (TGEV) (AAQ02624.1), feline infectious peritonitis virus (FIPV) (AZH81408.1), porcine epidemic diarrhea virus (PEDV) (AKP16765.2), equine coronavirus (BAJ52885.1), murine hepatitis virus (MHV) (BAA11889.1), bovine coronavirus (CCE89341.1), pangolin coronavirus (QIQ54048.1), bat coronavirus (AWV67072.1). Their nucleotide and S gene protein sequences were pooled using NCBI resource tools while analysis was done using EMBOSS needle, clustal W2 and clustal omega respectively.

Homology and phylogeny analysis of the S-protein genes in candidate zoonotic viruses: the identified spike gene protein sequences of animal

coronaviruses were retrieved from submitted protein entries in NCBI database, homology analysis of the sequences was compared using clustal omega, EMBOSS needle while phylogenetic trees was constructed using the neighbor-joining method by CLUSTAL X software.

SARS-CoV-2 sequence and phylogenetic analyses:

in total, we culled the respective genomic and protein data of eight [8] 2019-nCoV/SARS-CoV-2 clinical isolates from beta coronaviruses database in NCBI and these are: [1] MN908947; QHD43415.1 (China-Wuhan, December 2019), [2] LC522350;BBZ90167.1 (Philippines, January 2020), [3] LC523807;BCA37476.1 (Philippines, January 2020), [4] LC523808;BCA37477.1 (Philippines, January 2020); [5] MT308701;QIV64962.1 (Tunisia, April 2020) [6] MT308702;QIV64963.1 (USA, February 2020), [7] MT308703;QIV64975.1 (USA, April 2020) and [8] MT308704 (USA, April 2020). Whole-genome alignment and protein sequence identity calculation were performed using multiple sequence alignment in EMBL-EBI database with default parameters in clustal W2 and clustal omega respectively.

Results

Conserved domains in SARS-CoV-2: four out of 29 domain hits generated from 2019-nCoV/SARS-CoV-2 CDD query were selected based on the E-value scores (Table 1). These are: non-structural protein (nsp 11), coronavirus RPolN terminus, non-structural protein (nsp 13) and corona S2 super family.

Protein phylogeny assembly of SARS-CoV-2 isolates: protein sequence alignment analyses reveals the closest evolutionary conservation between 2019-nCoV/SARS-CoV-2 and pangolin S protein with 95.6% similarity and 92.1% identity while 46.8% similarity and 31.2% identity was observed between SARS-CoV-2 and bat S protein (Figure 2).

Evolutionary patterns from SARS-CoV-2 isolates: increased level of evolutionary divergence was observed in submitted entries of the recent SARS-CoV-2 genomic data during time of the study (entries from December 2019 till 4th April) as seen in Figure 3 and Figure 4, while evolutionary patterns observed between Wuhan SARS-CoV-2 data and other five geographical locations reveal trajectory of infection from reported source of outbreak.

Discussion

The region of 2019-nCoV domain which encodes nsp 11 spans from about 18046-19824bp. It was indicated in countering host innate antiviral response via inhibition of type I interferon (IFN) production using NendoU activity-dependent mechanisms in porcine reproductive syndrome viruses [16]. The nsp 11 is also associated with pathways such as programmed cell death evasion, mitogen-activated protein kinase signaling, histone-related, cell cycle and DNA replication and the ubiquitin-proteasome through RNA microarray analysis [17-20] and few nsp 11 inhibitors include papain-like proteinase (pIPRO) and 3C-like main protease-3CLpro [21]. Coronavirus RNA-directed RNA polymerase (RdRp) terminus covers the N-terminal region of the coronavirus. It spans from about 13480-14538bp in SARS-CoV-2 and its interaction with nsp3 has been indicated in viral replication especially during early onset of infection [22]. The inhibitors of coronavirus RdRp include ATP inhibitors with mfScores lower than 110 [21]. The nsp 13 is regarded as a highly conserved and multifunctional helicase unit and its spans from about 20662-21537 in the SARS-CoV-2 isolate [23].

They are SARS-CoV helicases that are chiefly concerned with RNA processing, DNA replication, recombination and repair, transcription and translation [24]. A few potential inhibitors of nsp13 have been identified [25,26] and they act by interfering with its unwinding and ATPase activities. The coronavirus S2 super family spans from 23546-25372 and forms the characteristic

'corona' after which the group is named. CoV diversity is reflected in the variable spike proteins (S proteins) and evolves into forms differing in receptor interactions and response to various environmental triggers of virus-cell membrane fusion [27]. The C-terminal (S2) domain directs ectodomain fusion of all CoVs spike proteins following receptor binding [28,29]. The level of interactions between the S protein and the virus receptor controls the host cell range [30]. A study showed a switch of species specificity via a mutant mouse hepatitis virus (MHV) construct which conferred horizontal gene transfer and ability to infect feline cells which were initially absent in wild MHV cells [30]. This was achieved via the substitution of the spike glycoprotein ectodomain. Another research [31] also indicated role of natural mutations in reactivity between the receptor binding domain of spike and cross-neutralization between palm civet coronavirus and SARS-CoVs.

Identification of the origin, natural host (s) and evolutionary pathway of viruses which causes pandemics is essential to understand molecular mechanism of their cross-species interactivity and implementation of a proper control measure [32]. Protein sequence alignment analyses reveals the closest evolutionary conservation between 2019-nCoV/SARS-CoV-2 and pangolin S protein with 95.6% similarity and 92.1% identity while 46.8% similarity and 31.2% identity was observed between SARS-CoV-2 and bat S protein (supplementary data). This finding therefore agrees with reports indicating pangolin as a more recent ancestor of SARS-CoV-2 than bats [33,34] which could have arisen as a result of recombination (chimera) or interactions between pangolin-CoV-like virus with a bat-CoV-RaTG13-like virus going by the homology and subclade of SARS-CoV-2 and pangolin S genes from bat S-gene seen in this study (Figure 2). Although, some computational analyses prediction of the improbability of direct binding between receptor binding domains (RBDs) in SARS-CoV-2 and ACE2 in humans suggests otherwise [35,36], studies have shown demonstrations of cross-species

interactivity through structural (*in-silico*), *in-vitro* and *in-vivo* mechanisms [31,37-39].

Series of *in-vivo* and *in-vitro* RNA recombination leading to vast genetic variability of positive strand RNA viruses has also been reported [13]. Domestication, consumption and wildlife activities which results in natural selection on a human or human-like ACE2 receptor [33,36] raises the possibility of SARS-CoV-2 emergence from pangolin. The receptor-binding domain (RBD) in the spike protein and functional polybasic (furin) cleavage site at the S1-S2 boundary [33,40] directs viral-host cell interactivity. Earlier reports of “no interactivity” between the S glycoprotein of bat-CoV with human receptor ACE2 (which is essential for cross-species transmission) was founded on significantly low levels of sequence similarity (79 to 80%) between bat-CoV and SARS-CoV in genes encoding the viral spike (S) glycoprotein [41]. However, the higher homology score (95.6%) in the receptor S-gene region obtained in this study suggests pangolin as the most likely intermediate host of SARS-CoV-2 than bats with 46.8% score (supplementary data).

Possible host-viral genetic recombination [42] could also account for the increased level of evolutionary divergence observed in submitted entries of the recent SARS-CoV-2 genomic data during time of the study. Evolutionary patterns observed between Wuhan SARS-CoV-2 data and other five geographical locations reveal trajectory of infection from reported source of outbreak. Mobility patterns of both humans and animals alike are major factors influencing zoonotic disease outbreak amongst populations [43]. This amongst others, necessitates the strict travel bans, laws and confinement strategies adopted in different countries to curb its spread. Surprisingly, genomic and protein data from Philippines suggests otherwise (Figure 3 and Figure 4). Despite the limited data used for SARS-CoV-2 genomic profiling in this study, we found viral subtype divergence (considering distance metrics of SARS-CoV-2 with entries) (Figure 3 and Figure 4) suggesting a population-specific post translational

modification which could have been influenced by genetic makeup. This is presumed based on subclades formed between protein sequence data from Philippines (BCA37476.1 and BCA37477.1) and another between China (QHD43415.1) and Philippines (BBZ90167.1) countries in the same continent. Also, empirical data points to genetic and epigenetic factors in SARS-CoVs evolution, incidence and infection rates amongst diverse populations and across different racial backgrounds [44].

Conclusion

Viral cellular mechanisms are vital factors necessary for replication during infection. Hence, identification of domains of viral entry and evasion of antiviral mechanisms in host cells is essential for development of effective therapeutic measures. Conserved domains that are vital targets sites for inhibition of SARS-CoV-2 viral entry and replication in host cells found in this study include nsp11, nsp 13, RdRp and corona super family while compounds such as RNA aptamers, ATP inhibitors, papain-like proteinase (pIPRO) and 3C-like main protease-3CLpro etc. are viable indicated inhibitors of these domains; also, understanding the evolutionary pathway of the novel coronavirus transmission will not only help combat the current pandemic but assist in mutation tracking for identifying future zoonotic coronaviruses threats. The phylogenetic analyses of candidate zoonotic coronavirus (S) gene with SARS-CoV-2 revealed pangolin as the most recent ancestor which formed a sub-clade with bat S-gene suggesting interspecies recombination of CoV in bats and pangolins. Evolutionary pattern observed between SARS-CoV-2 genomic data from source of outbreak with recent entries analyzed in this study showed relative trajectory course of infection from source to other places except protein data from Philippines suggesting earlier existence of SARS-CoV-2 which should be further investigated. Also, genomic and protein data revealed racial viral subtype divergence and rapid rate of mutation despite the novelty of the outbreak.

Precise dating of viral subtype divergence will enable researchers correlate divergence with epidemics and pandemics via viral sequence sampling for proper time-scale measurements of zoonotic threats in human populations. Therefore, there is an urgent need for large scale analysis and profiling of genetic data of SARS-CoV-2 in affected populations especially in Africa where there is paucity of genomic SARS-CoV data for effective therapeutic measures.

What is known about this topic

- Cell machinery for SARS-CoV-2 viral entry is through the binding of its surface proteins to host cell receptors-angiotensin converting enzyme 2 (ACE 2) in epithelial cells in hosts;
- Host factors accounts for a great deal of genome variability in viral recombinants which ranges from multi-resistance to evolutionary novelties;
- The level of interactions between the S protein and the virus receptor controls the host cell range.

What this study adds

- Conserved domains for inhibition of SARS-CoV-2 viral entry and replication in host cells found in this study include *nsp11*, *nsp13*, *RdRp* and corona super family;
- Compounds such as RNA aptamers, ATP inhibitors, papain-like proteinase (*pIPRO*) and 3C-like main protease-3CLpro etc. are viable indicated inhibitors of these domains;
- The phylogenetic analyses of candidate zoonotic coronavirus (*S*) gene with SARS-CoV-2 revealed pangolin as the most recent ancestor which formed a subclade with bat *S*-gene.

Competing interests

The authors declare no competing interests.

Authors' contributions

The conception and design was achieved by Pius A Okiki and Adejoke O Obajuluwa; acquisition of data, analysis and interpretation of data was done by Adejoke O Obajuluwa and Tiwalola M; Obajuluwa Adejoke O Obajuluwa and Olakunle B Afolabi drafted the article and revised it for important intellectual content; Pius A Okiki approved the final version to be published. All the authors have read and agreed to the final manuscript.

Table and figures

Table 1: conserved domains in SARS-CoV-2 genomic data

Figure 1: S, M, N and E regions of the MN908947 genomic data (NCBI database)

Figure 2: spike gene phylogeny of zoonotic CoVs with 2019 nCoV/SARS-CoV-2

Figure 3: genomic data phylogeny of selected SARS-CoV-2 entries

Figure 4: protein data phylogeny of selected SARS-CoV-2 entries

References

1. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. *Virology*. 2019;16(1): 69. [PubMed](#) | [Google Scholar](#)
2. Xia S, Yan L, Xu W, Agrawal AS, Algaissi A, Tseng CK *et al.* A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Sci Adv*. 2019;5(4): eaav4580. [PubMed](#) | [Google Scholar](#)
3. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery*. 2020;6: 14. [PubMed](#) | [Google Scholar](#)
4. Raoult D, Zumla A, Locatelli F, Ippolito G, Kroemer G. Coronavirus infections: epidemiological, clinical and immunological features and hypotheses. *Cell Stress*. 2020;4(4): 66-75. [PubMed](#) | [Google Scholar](#)

5. Qiu W, Chu C, Mao A, Wu J. The impacts on health, society and economy of SARS and H7N9 outbreaks in China: a case comparison study. *J Environ Public Health*. 2018 Jun 28;2018: 2710185. **PubMed** | **Google Scholar**
6. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Bio*. 2015;1282: 1-23. **PubMed** | **Google Scholar**
7. Raj VS, Mou H, Smits SL, Dekkers DH, Müller MA, Dijkman R *et al*. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*. 2013;495(7440): 251-4. **PubMed** | **Google Scholar**
8. Wang N, Shi X, Jiang L, Zhang S, Wang D, Tong P *et al*. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res*. 2013;23(8): 986-93. **PubMed** | **Google Scholar**
9. Tahir Ul Qamar M, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani S. Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. *J Transl Med*. 2019;17(1): 362. **PubMed** | **Google Scholar**
10. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X *et al*. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*. 2020;63(3): 457-460. **PubMed** | **Google Scholar**
11. Froissart R, Roze D, Uzest M, Galibert L, Blanc S, Michalakakis Y. Recombination every day: abundant recombination in a virus during a single multi-cellular host infection. *PLoS Biol*. 2005;3(3): e89. **PubMed** | **Google Scholar**
12. Li F. Structure, function and evolution of coronavirus spike proteins. *Annu Rev Virol*. 2016;3(1): 237-261. **PubMed** | **Google Scholar**
13. Burjaski JJ. Genetic recombination in plant-infecting messenger-sense RNA viruses: overview and research perspectives. *Fron Plant Sci*. 2013;4: 68. **PubMed** | **Google Scholar**
14. Domingo E. Mechanisms of viral emergence. *Vet Res*. 2010;41(6): 38. **PubMed** | **Google Scholar**
15. Irwin KK, Renzette N, Kowalik TF, Jensen JD. Antiviral drug resistance as an adaptive process. *Virus Evol*. 2016;2(1): vew014. **PubMed** | **Google Scholar**
16. Wang D, Chen J, Yu C, Zhu X, Xu S, Fang L *et al*. Porcine reproductive and respiratory syndrome virus nsp11 antagonizes type I interferon signaling by targeting IRF9. *J Virol*. 2019;93(15): e00623-19. **PubMed** | **Google Scholar**
17. Zhang M, Li X, Deng Z, Chen Z, Liu Y, Gao Y *et al*. Structural biology of the arterivirus nsp11 endoribonucleases. *J Virology*. 2017;91(1): e01309-16. **PubMed** | **Google Scholar**
18. He Q, Li Y, Zhou L, Ge XN, Guo X, Yang HC. Both Nsp1 beta and Nsp11 are responsible for differential TNF-alpha production induced by porcine reproductive and respiratory syndrome virus strains with different pathogenicity in vitro. *Virus Res*. 2015;201: 32-40. **PubMed** | **Google Scholar**
19. Sun Y, Li D, Giri S, Prasanth SG, Yoo D. Differential host cell gene expression and regulation of cell cycle progression by nonstructural protein 11 of porcine reproductive and respiratory syndrome virus. *Biomed Res Int*. 2014;2014: 430508. **PubMed** | **Google Scholar**
20. Yoo D, Song C, Sun Y, Du Y, Kim O, Liu HC. Modulation of host cell responses and evasion strategies for porcine reproductive and respiratory syndrome virus. *Virus Res*. 2010;154(1-2): 48-60. **PubMed** | **Google Scholar**
21. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y *et al*. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B*. 2020 May;10(5): 766-788. **PubMed** | **Google Scholar**
22. Tok TT, Tatar G. Structures and functions of coronavirus proteins: molecular modeling of viral nucleoprotein. *Int J Virol Infect Dis*. 2017;2(1): 001-007. **Google Scholar**

23. Jang KJ, Lee NR, Yeo WS, Jeong YJ, Kim DE. Isolation of inhibitory RNA aptamers against severe acute respiratory syndrome (SARS) coronavirus NTPase/Helicase. *Biochem Biophys Res Commun.* 2008;366(3):738-744. **PubMed** | **Google Scholar**
24. Snijder EJ, Decroly E, Ziebuhr J. The non-structural proteins directing coronavirus RNA synthesis and processing. *Adv Virus Res.* 2016;96: 59-126. **PubMed** | **Google Scholar**
25. Adedeji AO, Sarafianos SG. Antiviral drugs specific for coronaviruses in preclinical development. *Curr Opin Virol.* 2014;8: 45-53. **PubMed** | **Google Scholar**
26. Kim MK, Yu MS, Park HR, Kim KB, Lee C, Cho SY *et al.* 2, 6-Bisarylmethoxy-5-hydroxy chromones with antiviral activity against both hepatitis C virus (HCV) and SARS-associated coronavirus (SCV). *Eur J Med Chem.* 2011;46(11): 5698-5704. **PubMed** | **Google Scholar**
27. Heald-Sargent T, Gallagher T. Ready, set, fuse: the coronavirus spike protein and acquisition of fusion competence. *Viruses.* 2012;4(4): 557-580. **PubMed** | **Google Scholar**
28. Kirchdoerfer RN, Cottrell CA, Wang N, Pallesen J, Yassine HM, Turner HL *et al.* Pre-fusion structure of a human coronavirus spike protein. *Nature.* 2016;531(7592): 118-121. **PubMed** | **Google Scholar**
29. Tortorici MA, Walls AC, Lang Y, Wang C, Li Z, Koerhuis D *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nat Struct Mol Biol.* 2019;26(6): 481-489. **PubMed** | **Google Scholar**
30. Kuo L, Godeke GJ, Raamsman MJ, Masters PS, Rottier PJ. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J Virol.* 2000;74(3): 1393-1406. **PubMed** | **Google Scholar**
31. Liu L, Fang Q, Deng F, Wang H, Yi CE, Ba L *et al.* Natural mutations in the receptor binding domain of spike glycoprotein determine the reactivity of cross-neutralization between palm civet coronavirus and severe acute respiratory syndrome coronavirus. *J Virol.* 2007;81(9): 4694-4700. **PubMed** | **Google Scholar**
32. Zheng J. SARS-CoV-2: an emerging coronavirus that causes a global threat. *Int J Biol Sci.* 2020;16(10): 1678-1685. **PubMed** | **Google Scholar**
33. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr Biol.* 2020;30(7): 1346-1351. **PubMed** | **Google Scholar**
34. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou J *et al.* Isolation and characterization of 2019-nCoV-like coronavirus from Malayan pangolins. *bioRxiv.* 2020. **Google Scholar**
35. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J Virol.* 2020;94(7): e00127-20. **PubMed** | **Google Scholar**
36. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med.* 2020;26(4): 450-452. **PubMed** | **Google Scholar**
37. Ortega JT, Serrano ML, Pujol FH, Rangel HR. Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: an in silico analysis. *EXCLI J.* 2020;19: 410-417. **PubMed** | **Google Scholar**
38. Chen Y, Guo Y, Pan Y, Zhao ZJ. Structure analysis of the receptor binding of 2019-nCoV. *Biochem Biophys Res Commun.* 2020;525(1): 135-40. **PubMed** | **Google Scholar**
39. Sui J, Deming M, Rockx B, Liddington RC, Zhu QR, Baric RS *et al.* Effects of human anti-spike protein receptor binding domain antibodies on severe acute respiratory syndrome coronavirus neutralization escape and fitness. *J Virol.* 2014;88(23): 13769-13780. **PubMed** | **Google Scholar**

40. Lam TTY, Shum MHH, Zhu HC, Tong YG, Ni XB, Liao YS *et al.* Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China (Preprint). *bioRxiv.* 2020. **Google Scholar**
41. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science.* 2005 Sep 16;309(5742): 1864-8. **PubMed** | **Google Scholar**
42. Agudelo-Romero P, Carbonell P, Perez-Amador MA, Elena SF. Virus adaptation by manipulation of host's gene expression. *PLoS ONE.* 2008;3(6): e2397. **PubMed** | **Google Scholar**
43. Kluge HHP, Jakar Z, Bartovic J, D'Anna V, Severoni S. Refugee and migrant health in the COVID-19 response. *The Lancet.* 2020;395(10232): 1237-1239. **PubMed** | **Google Scholar**
44. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg Infect Dis.* 2020;26(7): 1470-1477. **PubMed** | **Google Scholar**

Table 1: conserved domains in SARS-CoV-2 genomic data

Domain hits	Accession no	Interval	e-value
NSP11	pfam06471	18046-19824	0e+00
Corona_RPol_N	pfam06478	13480-14538	0e+00
NSP13	pfam06460	20662-21537	0e+00
Corona (S2) super family	CI20218	23546-25372	0e+00

NSP-non-structural protein; RPol- RNA polymerase

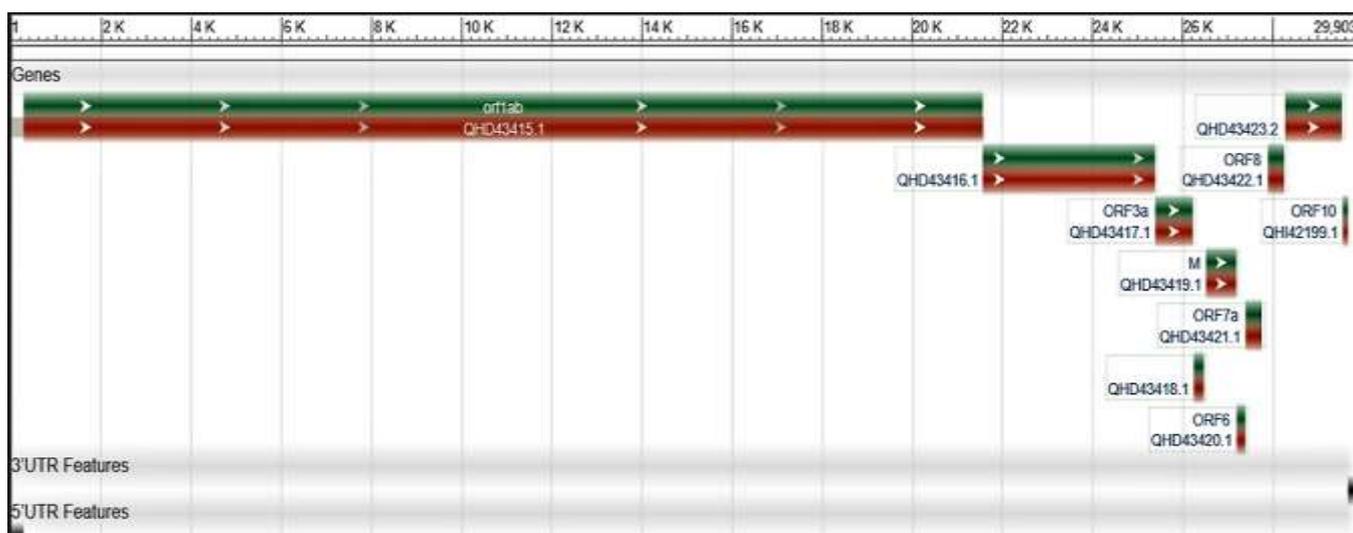


Figure 1: S, M, N and E regions of the MN908947 genomic data (NCBI database)

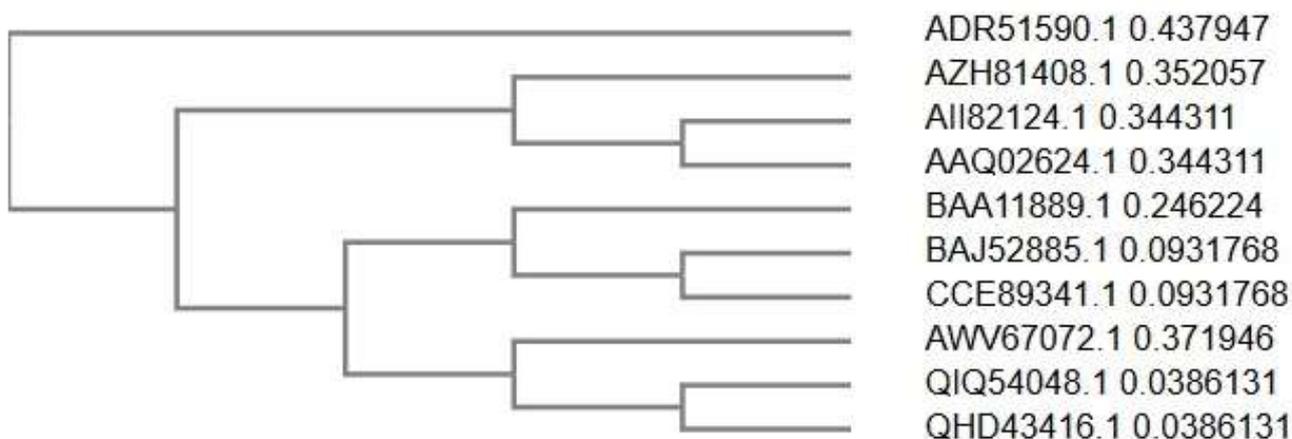


Figure 2: spike gene phylogeny of zoonotic CoVs with 2019 nCoV/SARS-CoV-2

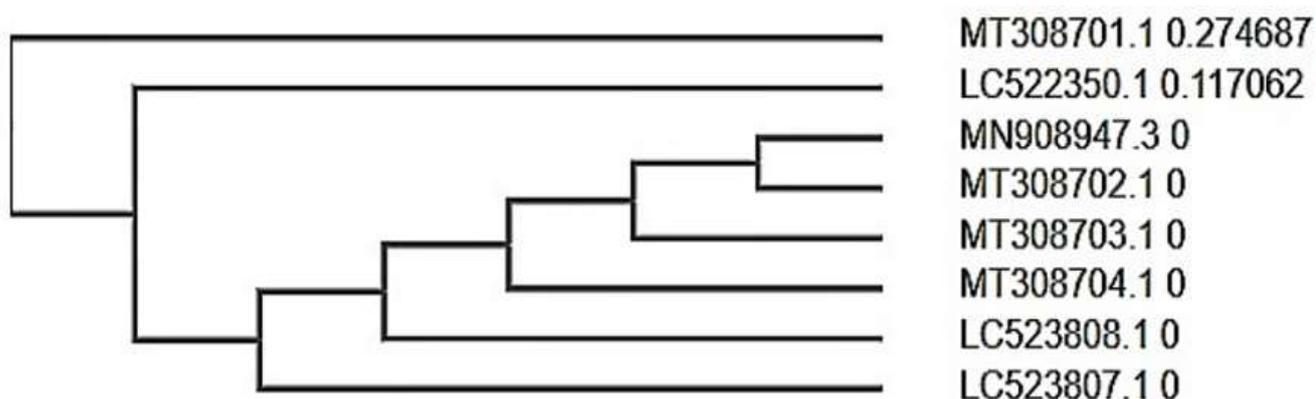


Figure 3: genomic data phylogeny of selected SARS-CoV-2 entries

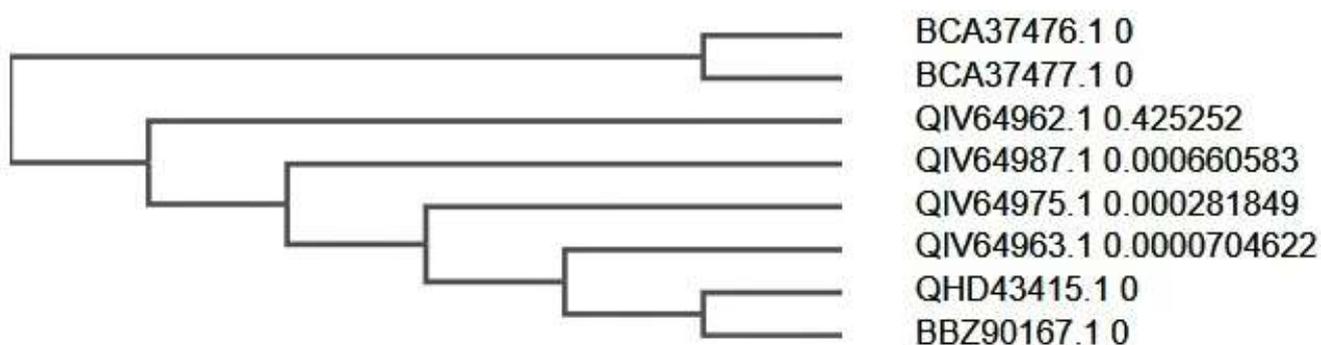


Figure 4: protein data phylogeny of selected SARS-CoV-2 entries