



BRIEF REPORT

Phylogenetic analysis of the betacoronavirus S1 subunit

[version 1; peer review: 1 approved with reservations]

Irina Zyrianova 

Institute for Innovative Biotechnologies in Animal Husbandry, the branch of L.K. Ernst Federal Science Center for Animal Husbandry, Moscow, Russian Federation

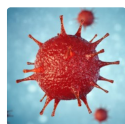
v1 First published: 03 Dec 2020, 9:1389
<https://doi.org/10.12688/f1000research.27681.1>
Latest published: 03 Dec 2020, 9:1389
<https://doi.org/10.12688/f1000research.27681.1>

Abstract

The ongoing pandemic outbreak of coronavirus disease 2019 (COVID-19) has been caused by the new betacoronavirus (BetaCoV) severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2). Together with other epidemic outbreaks of BetaCoV infectious diseases (Severe Acute Respiratory Syndrome (SARS) in 2002-2003 in China and Middle East Respiratory Syndrome (MERS) in 2012 in the Middle East, which have been caused by SARS-CoV and MERS-CoV, respectively), these events have generated interest in the coronaviruses (CoVs). Although many phylogenetic analyzes have been reported at a gene or protein level, there is no study as yet encompassing the many sequences publicly available for BetaCoVs, including those that have been manipulated in the lab. In this study, the phylogenetic analysis of 679 different S1 protein sequences of BetaCoVs from a total of 1595, which are publicly available in GenBank from the beginning of the pandemic event to April 2020, has been carried out. The S1 subunit is one part of the S (spike) protein, one of three CoV envelope proteins. The S1 subunit contains a host cell receptor binding domain. This domain is essential in the initiation of the infectious process. Therefore, its phylogenetic analysis is very important for studying CoV evolution. The phylogenetic analysis of BetaCoV S1 protein presented herein shows the evolutionary history of BetaCoVs from bovine CoV to SARS-CoV-2.

Keywords

S protein, S1 subunit, betacoronaviruses, bovine coronavirus, MERS-CoV, SARS-CoV, SARS-CoV-2, phylogenetic analysis



This article is included in the [Disease Outbreaks gateway](#).

Open Peer Review

Reviewer Status ?

Invited Reviewers

1

version 1

03 Dec 2020

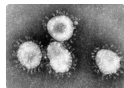
?

report

1. **Houcemeddine Othman** , Sydney

Brenner Institute for Molecular Bioscience,
Johannesburg, South Africa

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the **Coronavirus** collection.

Corresponding author: Irina Zyrianova (mirsimzyrianova@mail.ru)

Author roles: Zyrianova I: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: The study was funded by the Ministry of Science and Higher Education of the Russian Federation (№ 075-01250-20-01).

Copyright: © 2020 Zyrianova I. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Zyrianova I. **Phylogenetic analysis of the betacoronavirus S1 subunit [version 1; peer review: 1 approved with reservations]** F1000Research 2020, 9:1389 <https://doi.org/10.12688/f1000research.27681.1>

First published: 03 Dec 2020, 9:1389 <https://doi.org/10.12688/f1000research.27681.1>

Introduction

According to the [International Committee on Taxonomy of Viruses](#) (ICTV; 2019, Release #35), betacoronaviruses (BetaCoVs) have been classified as belonging to Riboviria realm, Orthornavirae kingdom, Pisuviricota phylum, Pisoniviricetes class, Nidovirales order, Cornidovirineae suborder, Coronaviridae family, Orthocoronavirinae subfamily, and Betacoronavirus genus; species/subspecies of BetaCoVs are listed in [Figure 1](#).

The ongoing pandemic outbreak of coronavirus disease 2019 (COVID-19) with pneumonia symptoms has been caused by a new BetaCoV, severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2; originally named as 2019-nCoV)^{1,2}. Two other BetaCoVs, SARS-CoV and MERS-CoV, have caused epidemic outbreaks of infectious diseases - Severe Acute Respiratory Syndrome (SARS) 2002–2003 in China and Middle East Respiratory Syndrome (MERS) in 2012 in the Middle East, respectively. All these outbreaks are severe or even fatal human diseases^{2,3}. The other three human BetaCoVs (human coronavirus OC43, NL63, and HKU1; HCoV-OC43, HCoV-NL63, and HCoV-HKU1), usually cause cold symptoms^{2,3}.

The rest of BetaCoVs primarily infect nonhuman mammals, among which bovine coronavirus (BCoV) is the most significant for the farming industry all over the world. Although, BCoV infected cattle have low mortality, they suffer from calf diarrhea (winter dysentery), respiratory symptoms, and substantial losses in milk and meat production⁴. Another BetaCoV, porcine hemagglutinating encephalomyelitis virus (PHEV), which is the causative agent of neurological and digestive disease in pigs, is also significant for farmers. Although it remains poorly studied because of its low clinical prevalence reported so far, it could lead to an animals' fatal end, causing significant harm to the swine industry⁵. Another BetaCoV is equine coronavirus (EqCoV), which causes diarrhea in foals and impacts the horse breeding industry⁶. There is also canine respiratory coronavirus (CRCoV), which is found in pet animals, associated with mild to severe canine infectious respiratory disease⁷.

Several small mammal BetaCoVs have been discovered to date. Among them are rodent coronavirus (RoCoV), rabbit coronavirus (RbCoV), and hedgehog coronavirus (HedCoV)⁸. There is also the more studied murine hepatitis virus (MHV), which causes hepatitis, enteritis, respiratory diseases, and encephalomyelitis in the central nervous system in mice and rats⁹. Furthermore, there are different bat coronaviruses (BatCoV), which have well-adapted hosts (different species of bats) in the natural environment. These BetaCoVs should be specially noted as they are suggested to be the origin of MERS-CoV, SARS-CoV, and SARS-CoV-2^{10,11}.

The S (spike) protein is one of three viral envelope proteins. It is considered a member of the class I viral membrane fusion proteins, including those from the influenza virus, human immunodeficiency virus (HIV), and Ebola virus. The S protein is involved in the initiation of the infectious process. It acts as an intermediary of viral and host cell membrane fusion and is a significant inducer of host immune responses^{12–15}. The S protein

assembles into trimers and folds so that it sticks out from the membrane surface to form spikes; hence its name: spike protein. The virion surface looks like a corona (Latin for crown) because of these spikes, and this feature became the reason for the name coronaviruses^{12–15}.

In most species of CoVs, the S protein is cleaved into two approximately equal size subunits, S1 and S2. The S1 subunit contains a host cell receptor binding domain (RBD). There are detected N-terminal domain (NTD) and C-terminal domain (CTD) in the S1 subunit, and one or both of which function as RBD. NTDs are responsible for binding the sugar (*O*-ac-Sia – 9-*O*-acetyl sialic acid) or the protein receptor CEACAM1 (the carcinoembryonic antigen-related adhesion molecule 1), and CTDs are responsible for recognizing protein receptors ACE2 (angiotensin-converting enzyme 2 – the zinc peptidase) and DPP4 (the dipeptidyl peptidase 4 – the serine peptidase), CD209L, and CD209 – the immunoglobulin-like cell adhesion molecule^{12–17}. The schematic structure of the Betacoronavirus S protein is shown in *Extended data*¹⁸.

The S1 subunit is the most divergent region of the S protein, and the S1 RBD is the principal determinant of species and tissue susceptible to infection¹². Therefore, its phylogenetic analysis is very important for studying coronavirus evolution. Although many phylogenetic analyzes have been reported for S or S1 on a genetic or protein level, no study had been made for all publicly available S1 protein sequences of all known BetaCoVs. In this study, the data of BetaCoV S1 protein sequence phylogenetic analysis has been presented. The S protein sequences used have been collected from GenBank before April 2020.

Methods

A total of 1595 S protein sequences, which are publicly available in GenBank from the beginning of the pandemic event to April 2020, have been used in this study. Some S protein sequences have been deduced from the corresponded nucleotide sequences using the GeneRunner program. Only 679 different of them have been implemented to the phylogenetic analysis since identical sequences do not contribute to phylogenetic relationships.

Identical S and S1 protein sequences have been found using the ClustalW option of the MEGA X (Version 10.0.5) program¹⁹ and excluded from the phylogenetic analysis. All identical S and S1 protein sequences are available as *Underlying data*²⁰. S1 ends for SARS-CoV-2 (-RRAR₆₈₅), SARS-CoV (-SLLR₆₆₇), MERS-CoV (-RSVR₇₅₁), HCoV-OC43 (-RRSR₇₅₇), HCoV-HKU1 (-RRKRR₇₆₀), MHV-A59 (-RRAHR₇₂₁), BCoV-Quebec (-RRSRR₇₆₈), BatCoV-HKU4 (-STFR₇₄₉), and BatCoV-HKU-5 (-RVRR₇₄₅) have been determined according to Millet and Whittaker (2014) and James *et al.* (2020)^{21,22}. The rest of the S1 subunit ends have been deduced using the S1 sequence alignment in ClustalW (see *Underlying data*²³).

Phylogenetic analysis has been performed with the MEGA X (Version 10.0.5) program¹⁹, using the Maximum Likelihood

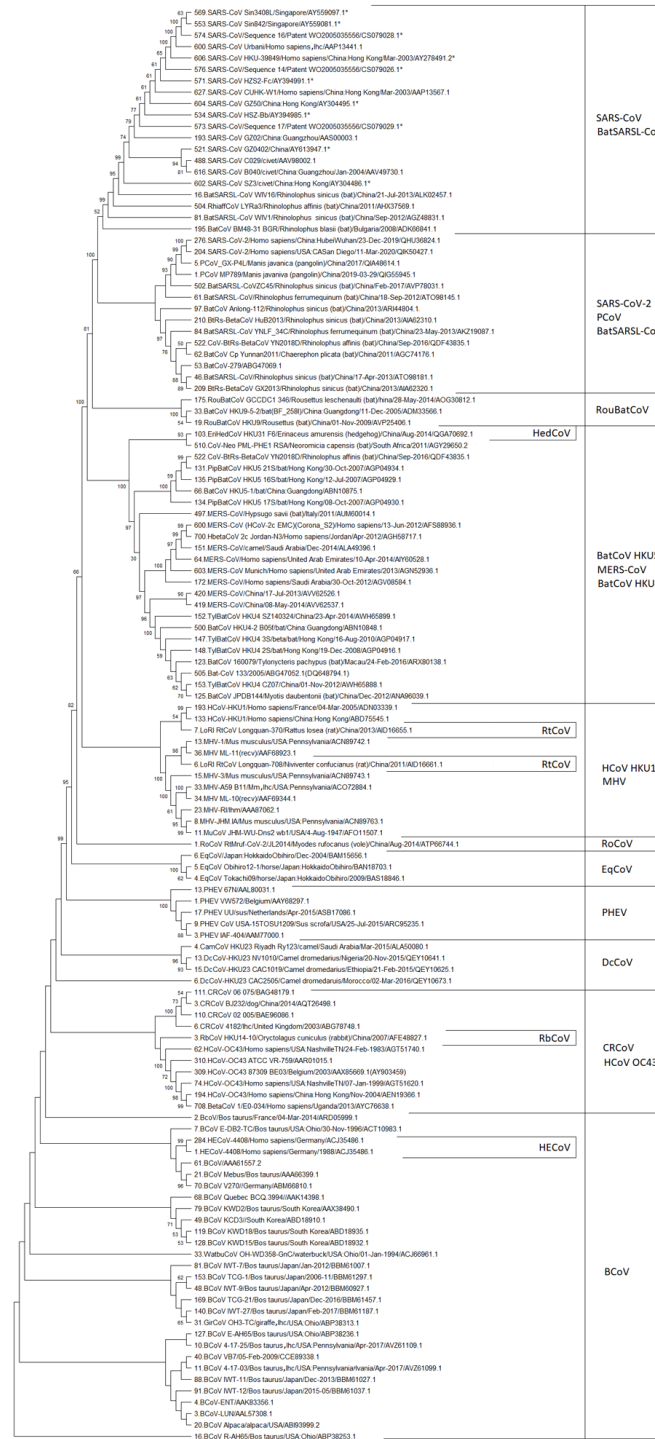


Figure 1. Phylogenetic analysis of the S1 protein sequence of betacoronaviruses. The following abbreviations are applied: lhc – the lab host cells, lhm – the lab host mouse; and SARS-CoV – Severe acute respiratory syndrome coronavirus, PCoV – Pangolin coronavirus, BatSARSL-CoV – Bat SARS-like coronavirus, BatCoV – Bat coronavirus, BetaCoV – Betacoronavirus, RoCoV – Rodent coronavirus, RtCoV – Rat coronavirus, RouBatCoV – Roussetts bat coronavirus, HCoV – Human coronavirus, HBetaCoV – Human betacoronavirus, HECoV – Human enteritis coronavirus, EriHedCoV – Erinaceus Hedgehog coronavirus, PipBatCoV – Pipistrellus bat coronavirus, HypBatCoV – Hypsugo bat coronavirus, MERS-CoV – Middle East respiratory syndrome coronavirus, TylBatCoV – Tylonycteris bat coronavirus, PhiaffCoV – Rhinolophus affinis coronavirus, CoV-Neo – Coronavirus Neoromocia, EqCoV – Equine coronavirus, MHV – Murine hepatitis virus, MuCoV – Murine coronavirus, PHEV – Porcine hemagglutinating encephalomyelitis virus, RbCoV – Rabbit coronavirus, DcCoV – Dromedary camel coronavirus, CamCoV – Camel coronavirus, CRCoV – Canine respiratory coronavirus, BCoV – Bovine coronavirus, WatbuCoV – Waterbuck coronavirus, GirCoV – Giraffe coronavirus. The stars designate protein sequences deduced from nucleotide sequences using the GeneRunner program. The numbers in front of sequence annotation are the unique sequence numbers for each S/S1 sequence in the batch for each BetaCoV species for more comfortable use. *Using data from 12–17.*

method and the JTT matrix-based model²⁴ with 1000 bootstrap replications and uniform rates among sites. The analysis of 679 different sequences of the S1 subunit has been implemented in two steps. In the first step, different phylogenetic trees have been constructed for each BetaCoV species/subspecies, or several BetaCoV species/subspecies have been combined into the one primary tree^{25–39}; except for human enteritis coronavirus (HECoV), for which all different protein sequences have been used in the summarised tree. RbCoV, RoCoV, rat (Rt) CoV, and HedCoV have been included together in the primary tree³⁴. Also, SARS-CoV-2 and pangolin coronavirus (PCoV) have been combined into the one primary tree³⁹. Bat CoVs have been divided into four groups by alignment; the primary tree has been constructed for each of them^{35–38}. In the second step, the sequences have been selected from primary trees, and a summarized tree has been constructed. As can be seen from each phylogenetic tree (see *Underlying data*^{25–39}) one or more sequences have been selected from each phylogenetic clade. The sequences have been chosen as follows: if the separated clade consists of several sequences, then the sequences that have been found closer to the branching point of the entire clade are selected; if the clade consists only of one sequence, this sequence is taken into the summary tree. The summary bootstrap consensus tree have been inferred from 1000 replicates. The percentage of replicate trees in which the associated taxa clustered together is shown next to the branches. Percentages $\geq 50\%$ are shown (Figure 1).

Results

Figure 1 shows that all S1 subunits of BetaCoVs species are originated from BCoV S1. The BCoV group also includes other ruminant BetaCoVs, which do not have an individual detached clade. There is an intermediate branch of human enteritis coronavirus among the BCoV group. It confirms the transmission of HECoV from bovine.

Furthermore, the phylogenetic clade of CRCoV and HCoV-OC43 has been separated from BCoVs, and then four clades have been detached one after another. These are Dromedary camel (Dc) CoV, PHEV, EqCoV, and RoCoV. They are followed by the clade consisting of MHV and HCoV-HKU1, which has several intermediates of RtCoVs.

After that, the group of MERS-CoV, BatCoV-HKU-4, and BatCoV-HRU-5 separate, and has an intermediate group of HedCoV. The MERS-CoV is already one of the particularly dangerous BetaCoVs for humans⁴⁰. This group is followed by the group of RouBatCoVs (Rousettus bat coronaviruses).

Finally, the SARS-CoV clade is formed. It is divided into two phylogenetic branches. One consists of SARS-CoVs and BatSARSL-CoVs (Bat SARS-like coronaviruses). Another consists of SARS-CoV-2s, PCoVs, and BatSARSL-CoVs. This clade could be named as that of the pandemic BetaCoVs, because of SARS-CoV-2.

Thus, we see in Figure 1 that the evolution of BetaCoV S1 proceeds from BCoV, which is not dangerous for humans, to

SARS-CoV and SARS-CoV-2, which are especially hazardous for humans.

Summary

The phylogenetic analysis carried out in this study has shown that the evolution of S1 of BetaCoVs begins from BCoV, which is not dangerous for humans, and then, passing through BetaCoVs of dogs (CRCoV), camels (DcCoV), pigs (PHEV), horses (EqCoV), rodents (RoCoV, MHV, RtCoV) and hedgehogs (HedCoV) leads to SARS-CoV and SARS-CoV-2, which are already particularly dangerous for humans. Therefore, we shouldn't underestimate the potential danger of BCoV.

Data availability

Underlying data

Figshare: 100% homology sequences of Betacoronaviruses S protein and S1 subunit, <https://doi.org/10.6084/m9.figshare.12962378.v4>²⁰.

Figshare: The S1 sequence alignment of Betacoronaviruses, <https://doi.org/10.6084/m9.figshare.13106894>²³.

Figshare: The phylogenetic analysis of the Bovine coronavirus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12956963.v5>²⁵.

Figshare: The phylogenetic analysis of the Human coronavirus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12957932.v4>²⁶.

Figshare: The phylogenetic analysis of the Severe Acute Respiratory Syndrome-related coronavirus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12957977.v5>²⁷.

Figshare: The phylogenetic analysis of the Middle East Respiratory Syndrome coronavirus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12957989.v4>²⁸.

Figshare: The phylogenetic analysis of the Murine hepatitis virus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958004.v4>²⁹.

Figshare: The phylogenetic analysis of the Canine respiratory coronavirus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958028.v4>³⁰.

Figshare: The phylogenetic analysis of the Porcine hemagglutinating encephalomyelitis virus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958091.v4>³¹.

Figshare: The phylogenetic analysis of the Equine coronavirus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958112.v4>³².

Figshare: The phylogenetic analysis of the Dromedary camel coronavirus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958136.v4>³³.

Figshare: The phylogenetic analysis of the small mammal BetaCoV coronavirus S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958172.v4>³⁴.

Figshare: The phylogenetic analysis of the Bat coronavirus (HKU3 group) S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958184.v4>³⁵.

Figshare: The phylogenetic analysis of the Bat coronavirus (HKU4 group) S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958193.v5>³⁶.

Figshare: The phylogenetic analysis of the Bat coronavirus (HKU5 group) S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958196.v5>³⁷.

Figshare: The phylogenetic analysis of the Bat coronavirus (HKU9,10 group) S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.12958208.v4>³⁸.

Figshare: The phylogenetic analysis of the Severe Acute Respiratory Syndrome-related coronavirus 2 S1 subunit protein sequence, <https://doi.org/10.6084/m9.figshare.13102889.v2>³⁹.

Extended data

Figshare: The schematic structure of the betacoronavirus S protein, <https://doi.org/10.6084/m9.figshare.12951413.v2>¹⁸.

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Acknowledgments

The author would like to thank Ilya Artemin – interpreter.

References

- Zhu N, Zhang D, Wang W, et al.: **A novel coronavirus from patients with pneumonia in China, 2019.** *N Engl J Med.* 2020; **382**(8): 727–733.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nadeem MS, Zamzami MA, Choudhry H, et al.: **Origin, Potential Therapeutic Target and Treatment for Coronavirus Disease (COVID-19).** *Pathogens.* 2020; **9**(4): 307.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ye ZW, Yuan S, Yuen KS, et al.: **Zoonotic origins of human coronaviruses.** *Int J Biol Sci.* 2020; **16**(10): 1686–1697.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kanno T, Ishihara R, Hatama S, et al.: **A long-term animal experiment indicating persistent infection of bovine coronavirus in cattle.** *J Vet Med Sci.* 2018; **80**(7): 1134–1137.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mora-Díaz JC, Piñeyro PE, Houston E, et al.: **Porcine Hemagglutinating Encephalomyelitis Virus: A Review.** *Front Vet Sci.* 2019; **6**: 53.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nemoto M, Oue Y, Higuchi T, et al.: **Low prevalence of equine coronavirus in foals in the largest thoroughbred horse breeding region of Japan, 2012–2014.** *Acta Vet Scand.* 2015; **57**(1): 53.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buonavoglia C, Martella V: **Canine respiratory viruses.** *Vet Res.* 2007; **38**(2): 355–73.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Monchatre-Leroy E, Boué F, Boucher JM, et al.: **Identification of Alpha and Beta Coronavirus in Wildlife Species in France: Bats, Rodents, Rabbits, and Hedgehogs.** *Viruses.* 2017; **9**(12): 364.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chakravarty D, Saadi F, Kundu S, et al.: **CD4 Deficiency Causes Poliomyelitis and Axonal Blebbing in Murine Coronavirus-Induced Neuroinflammation.** *J Virol.* 2020; **94**(14): e00548–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cui J, Li F, Shi ZL: **Origin and evolution of pathogenic coronaviruses.** *Nat Rev Microbiol.* 2019; **17**(3): 181–192.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zhou P, Yang XL, Wang XG, et al.: **A pneumonia outbreak associated with a new coronavirus of probable bat origin.** *Nature.* 2020; **579**(7798): 270–273.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Masters PS: **The Molecular Biology of Coronaviruses.** *Adv Virus Res.* 2006; **66**: 193–292.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lu G, Wang Q, Gao GF: **Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond.** *Trends Microbiol.* 2015; **23**(8): 468–478.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li F: **Structure, Function, and Evolution of Coronavirus Spike Proteins.** *Annu Rev Virol.* 2016; **3**(1): 237–261.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Satarkar S, Nampoothiri M: **Structural Proteins in Severe Acute Respiratory Syndrome Coronavirus-2.** *Arch Med Res.* 2020; **51**(6): 482–491.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Amraie R, Napoleon MA, Yin W, et al.: **CD209/L-SIGN and CD209/DC-SIGN act as receptors for SARS-CoV-2 and are differentially expressed in lung and kidney epithelial and endothelial cells.** *bioRxiv.* 2020; 2020.06.22.165803.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Naskalska A, Dabrowska A, Szczepanski A, et al.: **Membrane Protein of Human Coronavirus NL63 Is Responsible for Interaction with the Adhesion Receptor.** *J Virol.* 2019; **93**(19): e00355–19.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zyrianova I: **The schematic structure of the betacoronavirus S protein.** *figshare.* Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12951413.v2>
- Kumar S, Stecher G, Li M, et al.: **MEGA X: molecular evolutionary genetics analysis across computing platforms.** *Mol Biol Evol.* 2018; **35**(6): 1547–1549.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zyrianova I: **100% homology sequences of Betacoronaviruses S protein and S1 subunit.** *figshare.* Dataset. 2020.
<http://www.doi.org/10.6084/m9.figshare.12962378.v4>
- Millet JK, Whittaker GR: **Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis.** *Virus Res.* 2015; **202**: 120–134.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jaimes JA, André NM, Chappie JS, et al.: **Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop.** *J Mol Biol.* 2020; **432**(10): 3309–3325.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zyrianova I: **The S1 sequence alignment of Betacoronaviruses.** *figshare.* Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.13106894>
- Jones DT, Taylor WR, Thornton JM: **The rapid generation of mutation data matrices from protein sequences.** *Bioinformatics.* 1992; **8**: 275–282.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zyrianova I: **The phylogenetic analysis of the Bovine coronavirus S1 subunit protein sequence.** *figshare.* Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12956963.v5>
- Zyrianova I: **The phylogenetic analysis of the Human coronavirus S1 subunit protein sequence.** *figshare.* Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12957932.v4>
- Zyrianova I: **The phylogenetic analysis of the Severe Acute Respiratory Syndrome-related coronavirus S1 subunit protein sequence.** *figshare.* Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12957977.v5>
- Zyrianova I: **The phylogenetic analysis of the Middle East Respiratory Syndrome coronavirus S1 subunit protein sequence.** *figshare.* Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12957989.v4>
- Zyrianova I: **The phylogenetic analysis of the Murine hepatitis virus S1**

- subunit protein sequence. *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12958004.v4>
30. Zyrianova I: **The phylogenetic analysis of the Canine respiratory coronavirus S1 subunit protein sequence.** *figshare*. Figure. 2020.
<https://doi.org/10.6084/m9.figshare.12958028.v4>
 31. Zyrianova I: **The phylogenetic analysis of the Porcine hemagglutinating encephalomyelitis virus S1 subunit protein sequence.** *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12958091.v4>
 32. Zyrianova I: **The phylogenetic analysis of the Equine coronavirus S1 subunit protein sequence.** *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12958112.v4>
 33. Zyrianova I: **The phylogenetic analysis of the Dromedary camel coronavirus S1 subunit protein sequence.** *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12958136.v4>
 34. Zyrianova I: **The phylogenetic analysis of the small mammal BetaCoV coronavirus S1 subunit protein sequence.** *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12958172.v4>
 35. Zyrianova I: **The phylogenetic analysis of the Bat coronavirus (HKU3 group) S1 subunit protein sequence.** *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12958184.v4>
 36. Zyrianova I: **The phylogenetic analysis of the Bat coronavirus (HKU4 group) S1 subunit protein sequence.** *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12958193.v5>
 37. Zyrianova I: **The phylogenetic analysis of the Bat coronavirus (HKU5 group) S1 subunit protein sequence.** *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12958196.v5>
 38. Zyrianova I: **The phylogenetic analysis of the Bat coronavirus (HKU9,10 group) S1 subunit protein sequence.** *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12958208.v4>
 39. Zyrianova I: **The phylogenetic analysis of the Severe Acute Respiratory Syndrome-related coronavirus 2 S1 subunit protein sequence.** *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.13102889.v2>
 40. Yuan L, Tang Q, Cheng T, *et al.*: **Animal models for emerging coronavirus: progress and new insights.** *Emerg Microbes Infect.* 2020; 9(1): 949–961.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 02 February 2021

<https://doi.org/10.5256/f1000research.30599.r77185>

© 2021 Othman H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Houcemeddine Othman

Sydney Brenner Institute for Molecular Bioscience, Johannesburg, South Africa

The paper by Irina Zyrianova provides phylogenetic evidence about the evolution of SARS-CoV-2 from bovine coronavirus. Although the topic is interesting and could lead to establishing better monitoring strategies for beta-coronaviruses, the author could have discussed the results in a more elaborated way. The experimental protocol is not clear and many ambiguities need to be clarified to the reader.

Also, the phylogenetic analysis is only reported for the spike protein. It is known that the evolution rate of the protein S is higher than the entire genome. In this regard, getting general conclusions about the evolutionary history would not be appropriate.

Other minor comments are described below:

Abstract:

- *"including these that have been manipulated in the lab"* - It is not clear what the author means exactly by this.
- *"there is no study as yet encompassing the many sequences publicly available for BetaCoVs"* - I don't think this is accurate. Maybe the author could soften the claim a bit or be more specific given that other works tried to analyze the evolutionary history of SARS-CoV-2^{1,2}.
- *"The S1 subunit is one part of the S (spike) protein, one of three CoV envelope proteins."* - perhaps reformulate?

Figure 1: It should be referenced in the Results section and not the Introduction. A table can be useful to show the species/subspecies of BetaCovs.

Paragraph 4, Introduction: I am not sure why the author mentioned the hepatitis virus.

Paragraph 5, Introduction: A reference is needed for the first sentence.

Paragraph 6, Introduction: "one or both of which function as RBD", NTD and CTD are independent of RBD. Check the original paper describing the structure of the spike protein³.

Paragraph 1, Methods:

- "since identical sequences do not contribute to phylogenetic relationships", maybe "Do not contribute to adding information about the evolutionary relationship of the taxa".
- Can the author please specify if the set of sequences are nucleotide or protein sequences? It is better to study the phylogeny using nucleotide sequences, as it is more capable to discriminate the evolutionary events at short periods. Although it should use one of the alignment algorithms that allows for taking into consideration the coding nature of these sequences to reduce the alignment errors.

Paragraph 2, Methods:

- It is not clear what the author did here. Is it a redundancy removal stage? If so wasn't that done in the previous paragraph?
- How could ClustalW be used to do that?
- What is the meaning of "deducing the ends" and what is the reason behind that?

I do not recommend using two stages to build the tree. I can understand that this has been done to ease the figure, but there are many software and packages available to collapse the clades without losing too much information. Also, I noticed that some branches lack the bootstrap values including the divergence between the BCoV and CrCoV and HCoV OC43.

References

1. Li T, Liu D, Yang Y, Guo J, et al.: Phylogenetic supertree reveals detailed evolution of SARS-CoV-2. *Scientific Reports*. 2020; **10** (1). [Publisher Full Text](#)
2. Jaimes J, André N, Chappie J, Millet J, et al.: Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. *Journal of Molecular Biology*. 2020; **432** (10): 3309-3325 [Publisher Full Text](#)
3. Wrapp D, Wang N, Corbett K, Goldsmith J, et al.: Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020; **367** (6483): 1260-1263 [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research