ONOMY Science

Scientific Journal of Pharmacy & Pharmaceutical Sciences
ISSN 2688-5247

Short communication Open Access

The dark side of COVID-19

Antoni Romeu* and Enric Ollé

[1] Department of Biochemistry and Biotechnology, Rovira i Virgili University, E-43007 Tarragona, Spain.

***Corresponding Author: Antoni Romeu**, Department of Biochemistry and Biotechnology, Rovira i Virgili University, E-43007 Tarragona, Spain.; Email: antonioramon.romeu@iubilo.urv.cat

Citation: The dark side of COVID-19, Sci J Phar and Pharmaceu Sci. 2020; 2(1): 01-08.

Submitted: 22 October 2020; **Approved:** 26 October 2020; **Published:** 28 October 2020

ABSTRACT

The origin of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is unknown. Studying it requires also identifying the origin of the most related coronaviruses. For this purpose, in full genomic analysis, we found three genomic fingerprints of the COVID-19 related coronavirus: at the beginning of the Orf1ab RNA replicase gene ; at S gene; and the NS8 gene itself, at the end of the genome. Only the Bat SARS-like coronavirus (Bat-SL-CoV, isolates ZXC21 and ZC45) genomes matched the COVID-19 genome extreme fingerprints. The S gene fingerprint was unmatched: it is the key to the SARS-CoV-2 origin. Further phylogenetic analysis showed an orthology relationship between BatCoV-RaTG13, SARS-CoV-2, Bat-SL-CoV and Panglolin-CoV (isolate MP789). This phylogenetic relationship was broken when the phylogenetic analysis was based on the spike glycoprotein sequences, with a high divergence of BatCoV-RaTG13 and SARS- CoV-2. However, the polybasic furin cleavage site in the SARS-CoV-2 spike glycoprotein made the difference from SARS-CoV-2 to BatCoV-RaTG13. Unfortunately, there is only one freely available genome of BatCoV-RaTG13 in GenBank, both to statistically validate that 96% and to explore the remaining 4%. Because that polybasic furin cleavage site is responsible for the high SARS-CoV-2 infectivity and transmissibility, the availability of multiple BatCoV-RaTG13 complete genomes for comparative analysis would provide new perspective. Until this happens, the origin of SARS-CoV-2 is the dark side of COVID-19.

KEY WORDS: COVID-19, SARS-CoV-2, BatCoV-RaTG13, Coronavirus, Bioinformatics

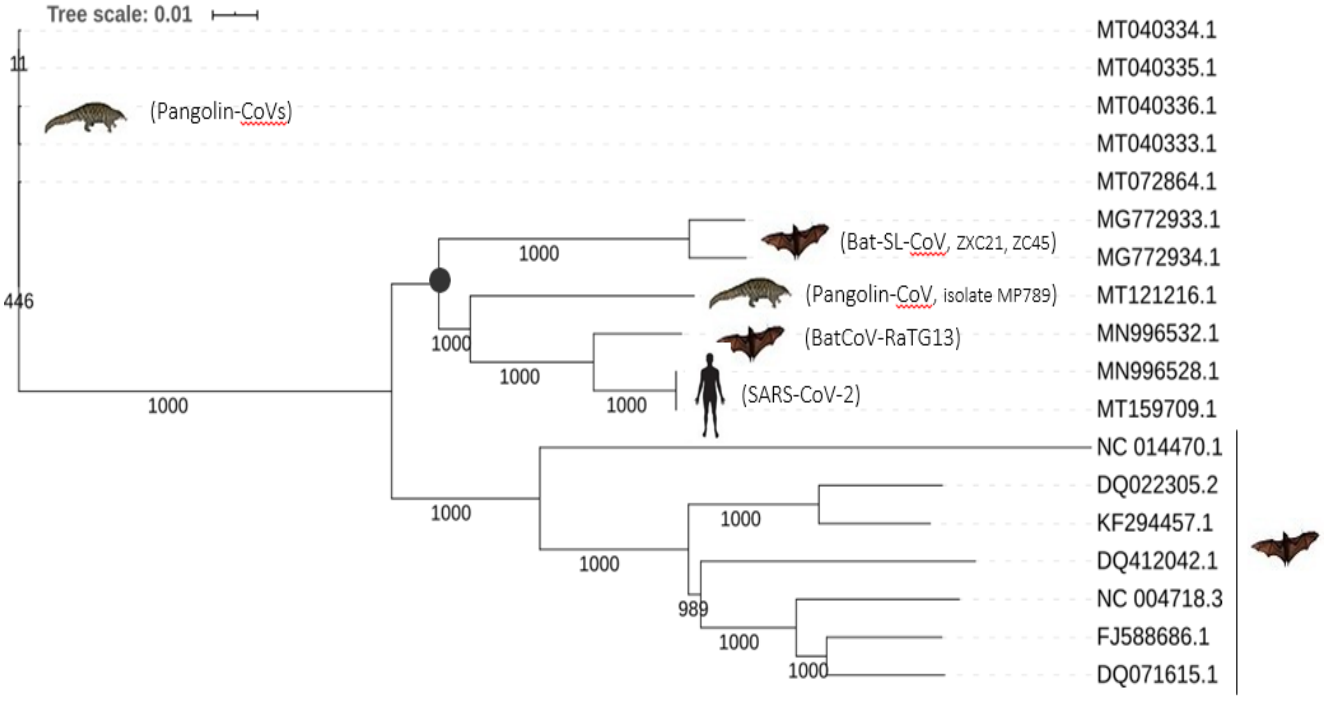
The origin of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the coronavirus disease 2019 (COVID-19), is unknown. Its origin is linked to that of the most SARS-CoV-2 related coronaviruses. We addressed the issue through a bioinformatic approach, providing the following to previous knowledge.

- First, using the Basic Local Alignment Search Tool (BLAST) of the National Center for Biotechnology Information (NCBI), there were clearly identified three genomic fingerprints of the COVID-19 related coronavirus: (i) the N-terminal Macro domain (ATP binding module) of the Orf1ab RNA replicase gene (based on SARS-CoV-2 genome, GenBank acession number MN996528.1, coordinates: 1940-3955); (ii) the N-terminal domain and receptor binding domain (RBD) of the spike glycoprotein, S gene (genome coordinates, 21563-22963); and (iii) the NS8 gene itself (genome coordinates, 27912-28256). Only the sequences of the Bat SARS-like coronavirus (Bat-SL-CoV) (isolates ZXC21 and ZC45, GenBank accession numbers MG772934.1 and MG772933.1, respectively) gave a perfect match with the extremes genome fingerprints. On the other hand, the spike glycoprotein fingerprint was unmatched: it is the key to the SARS-CoV-2 origin.
- Second, the phylogenetic analysis based on complete genomes, showed that Bat- SL-CoV (ZXC21 and ZC45), Pangolin-CoV (isolate MP789, GenBank accession number MT121216.1), BatCoV-RaTG13 (GenBank accession number MN996532.1) and SARS-CoV-2 were orthologous. That is, they originated from

a common ancestor coronavirus species, which were separated from each other after a speciation events. It is worth noting that not all pangolin coronaviruses are the same. Only the Panglolin-CoV (MP789) belonged to that group of orthologous coronavirus, other pangolin coronavirus were phylogenetically more distant (Figure 1).

Figure 1. Phylogenetic tree based on coronavirus complete genome

Phylogenetic tree based on a multialignment of complete genomes of selected coronaviruses. The sample includes the available pangolin coronaviruses, a selection of bat coronaviruses extracted from the literature (1,15,16), and two SARS-CoV-2 genomes as representatives of the NCBI "Severe acute respiratory syndrome coronavirus 2 (taxid:2697049)" taxonomic group. The phylogenetic tree was constructed with the Neighbor Joining method of the Clustal Omega (v.1.2.4) software package using default parameters (17). Assesed clustering strength was calculated by bootstrap using 1000 replicates. Tree scale bar stands for the evolutionary distance, based on genome multialigment. The black point depicts the ancestral coronavirus species from which the group of homologous sequences were separated from each other after a speciation event. GenBank accession number of the complete genomes and the coronavirus were the following (in the same arrangement as in the phylogenetic tree): MT040333.1 to MT040336.1, Pangolin coronavirus; MG772933.1 and MG772934.1, Bat SARS-like coronavirus; MT121216.1 (isolate MP789) Pangolin coronavirus; MN996532.1, Bat coronavirus RaTG13; MN996528.1 and MT159709.1, SARS-CoV-2; NC_014470.1, Bat coronavirus BM48-31/BGR/2008; DQ022305.2, Bat SARS coronavirus HKU3-1; KF294457.1,Bat SARS-like coronavirus; DQ412042.1, Bat SARS CoV Rf1/2004; NC_004718.3, SARS coronavirus Tor2; FJ588686.1, SARS coronavirus Rs_672/2006; DQ071615.1, Bat SARS CoV Rp3/2004.

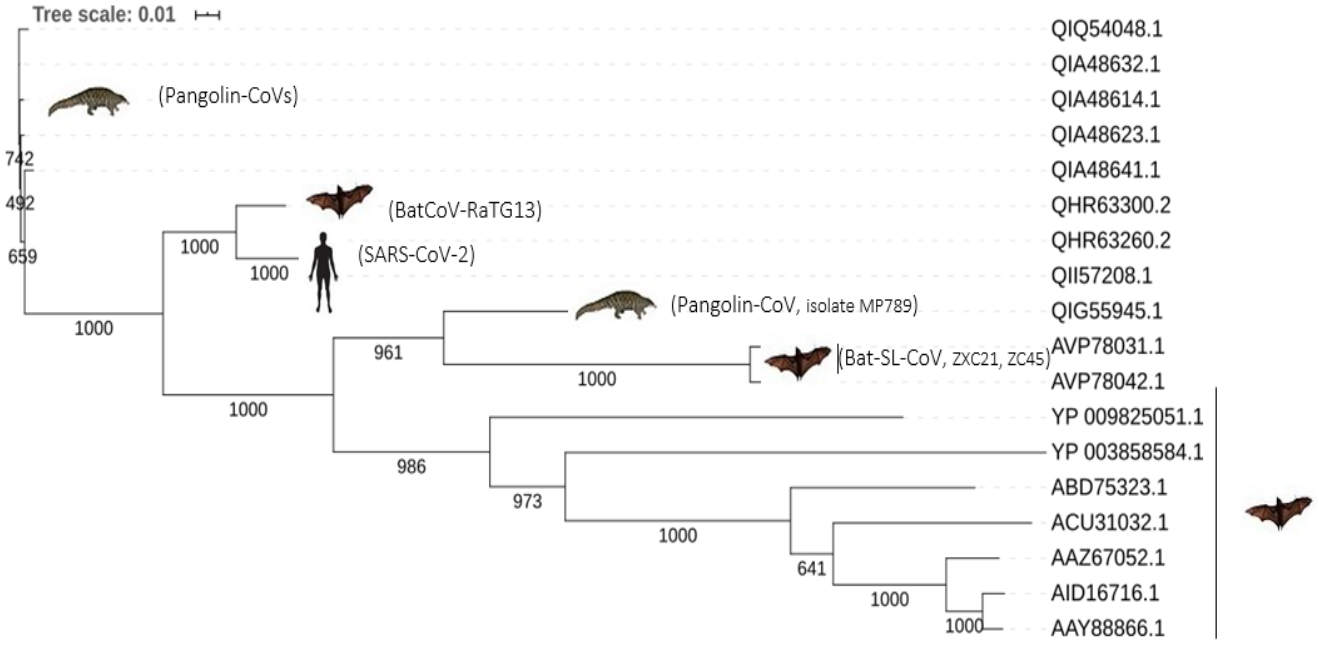


•Third, the phylogenetic relationship of orthologous was broken when the phylogenetic analysis was based on the spike glycoprotein sequences (specifically, from the N-terminal domain to the RBD, inclusive). In this new analysis, using the protein sequences from the same coronaviruses used in the analysis shown in Figure 1, the sequences of BatCoV-RaTG13 and SARS-CoV-2 diverged significantly from those of Bat-SL-CoV (ZXC21 and ZC45) and Pangolin-CoV (MP789). Furthermore, the last two were clearly grouped with the others bat coronaviruses (Figure 2).

Figure 2. Phylogenetic tree based on coronavirus spike glycoprotein

Phylogenetic tree based on the multialignment of the most variable region of the coronavirus spike glycoprotein: from the N-terminal domain to the RBD, inclusive. In this phylogenetic analysis, we have tried to use the sequences of the same coronaviruses of Figure 1, with the intention of comparing the phylogenetic behavior, when it is based on the complete genome and when it is based on this strategic region of the genome. RBD position was based on (18). The phylogenetic tree was constructed with the Neighbor Joining method of the Clustal Omega (v.1.2.4) software package, using default parameters (17). Assesed clustering strength was calculated by bootstrap using 1000 replicates. Tree scale bar stands for the evolutionary distance, based on sequence multialigmnet.GenBank accession number of the spike glycoprotein and the coronavirus were the following (in the same arrangement as in the phylogenetic tree): QIA48641.1, QIA48632.1, QIA48614.1, QIA48623.1, QIQ54048.1, Pangolin coronavirus; QHR63300.2, Bat coronavirus RaTG13; QHR63260.2, QII57208.1, SARS-CoV-2; QIG55945.1,

Pangolincoronavirus; AVP78031.1, AVP78042.1, Bat SARS-like coronavirus; YP_009825051.1, SARS coronavirus Tor2; YP_003858584.1, Bat coronavirus BM48-31/BGR/2008; ABD75323.1, Bat SARS CoV Rf1/2004; ACU31032.1, SARS coronavirus Rs_672/2006; AAZ67052.1, Bat SARS CoV Rp3/2004; AID16716.1, Bat SARS-like coronavirus; AAY88866.1, Bat SARS coronavirus HKU3-1.



BatCoV-RaTG13 has been considered likely to be the direct progenitor of SARS-CoV-2 (1), despite contact between humans and bats is limited (2). But, how and when did the Bat-CoV-RaTG13 appear? The BatCoV-RaTG13 genome was identified and sequenced by Zheng-Li Shi and coworkers, 2020 (1), within the framework of the study of a "new coronavirus", which caused the epidemic of acute respiratory syndrome in humans in Wuhan, China (December 2019). Literally: "We found that a short region of RNA- dependent RNA polymerase (RdRp) from a bat coronavirus (BatCoV RaTG13) —which was previously detected in Rhinolophus affinis from Yunnan province— showed high sequence identity to 2019-nCoV (SARS-CoV-2)". Then, Zheng-Li Shi and coworkers, 2020 (1), carried out full-length BatCoV-RaTG13 and SARS-CoV-2 genome sequencing, with an overall genome sequence identity of 96.2%. BatCoV-RaTG13 complete genome sequence was submitted to GenBank by Z.-L. Shi and coworkers on 27-JAN-2020 (MN996532.1). In the GenBank file, it appears as isolation source, fecal swab; host, Rhinolophus affinis; country, China; and collection date, 24-Jul-2013. From this available information, the origin of BatCoV-RaTG13 was prior to 2013. Accordingly, at least since 2013, it must be considered as an active bat coronavirus.

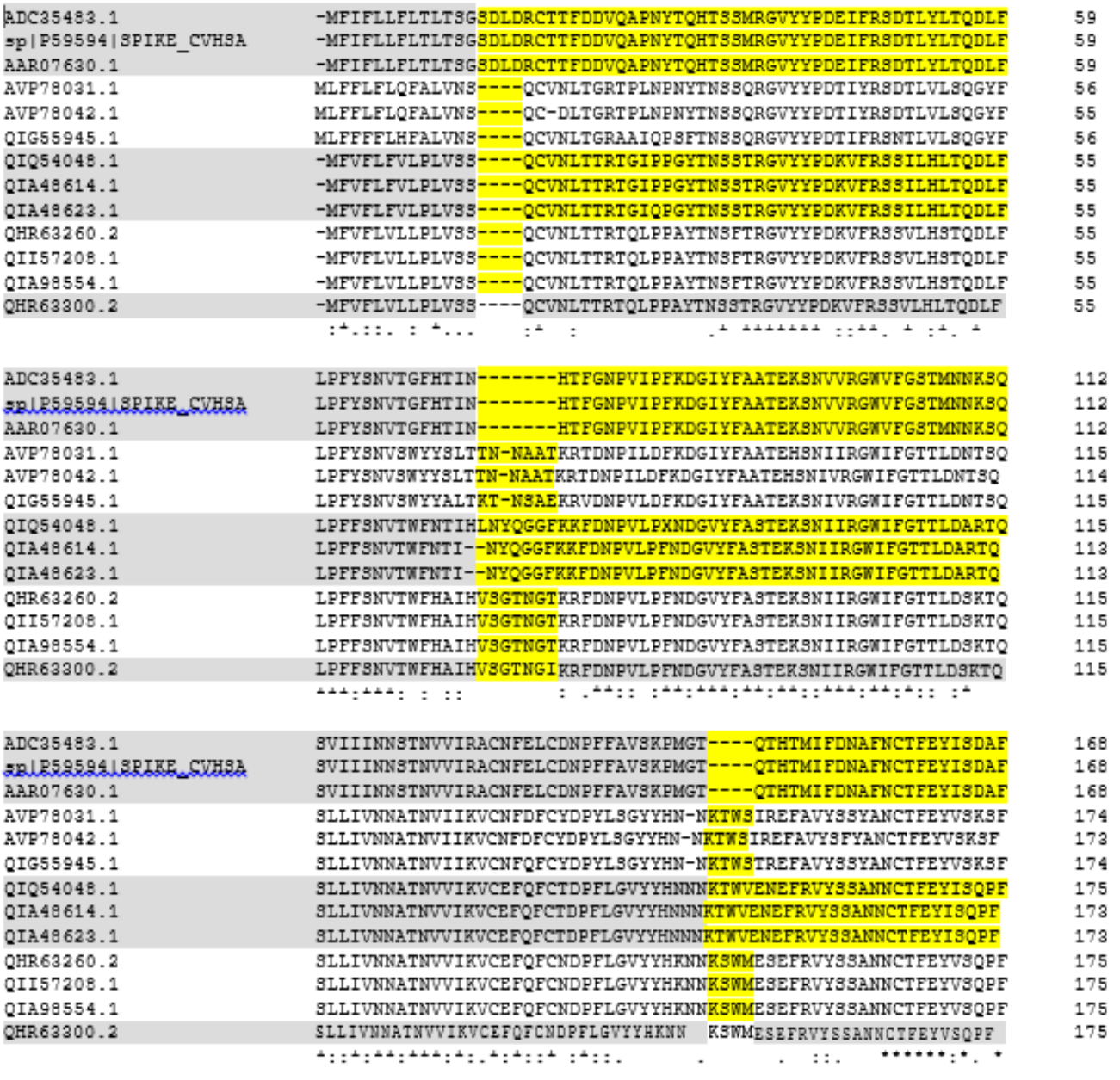
Regarding the spike glycoprotein phylogenetic divergence of BatCoV-RaTG13 and SARS-CoV-2, the RBD region of the spike glycoprotein is described as the most variable region of the coronavirus genomes (3). However, as it is shown in the multialignment (Figure 3), the N-terminal domain and RBD region is highly conserved, and many amino acid residues were strictly conserved. So that, according to Motoo Kimura's neutral theory of molecular evolution (4), in this critical region of the coronavirus genomes the balance between the random drift and the functional constraints would be enormously stressed. Nevertheless, the SARS-CoV-2 sipke glycoprotein appears to be optimized for binding to the human receptor ACE2 (3,5). Of course it has. Through The RBD of the spike glycoprotein attaches the virion to the cell membrane by interacting with host receptor, initiating the infection (5,6), and entry into human cells (7).

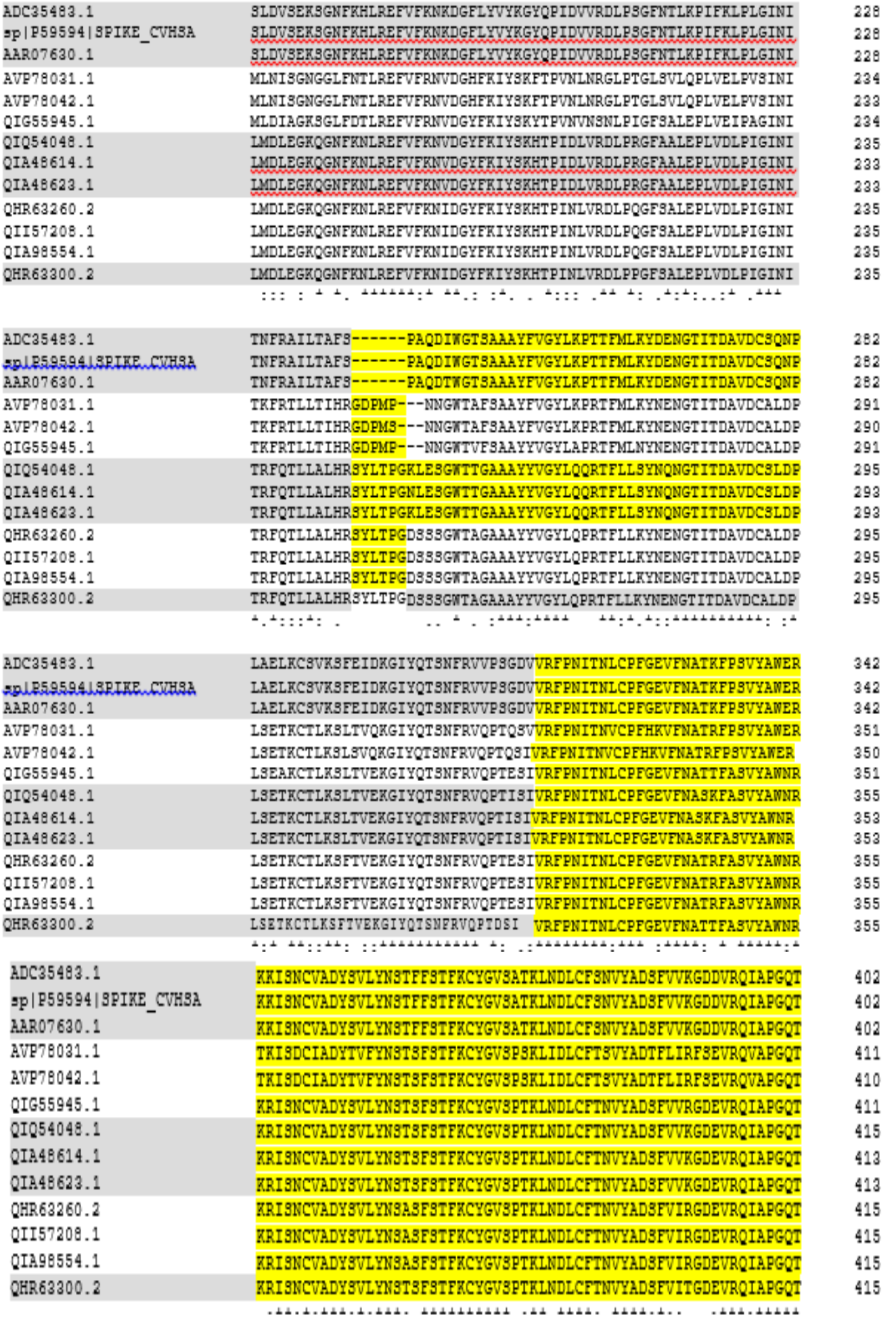
What evolutionary event could explain this RBD extraordinary optimization for human infection? A possible recombination event between BatCoV-RaTG13 and some bat coronavirus, or a horizontal gene transfer that had replaced variable regions in BatCoV- RaTG13 S gene, could be considered (8,9). But, either in recombination or a horizontal transfer, there always had to be a donor, and we could not identify any donor by analysing the entire NCBI nucleotide collection. This reasoning agrees with Zheng-Li Shi and coworkers, 2020 (1), literally: "Using the

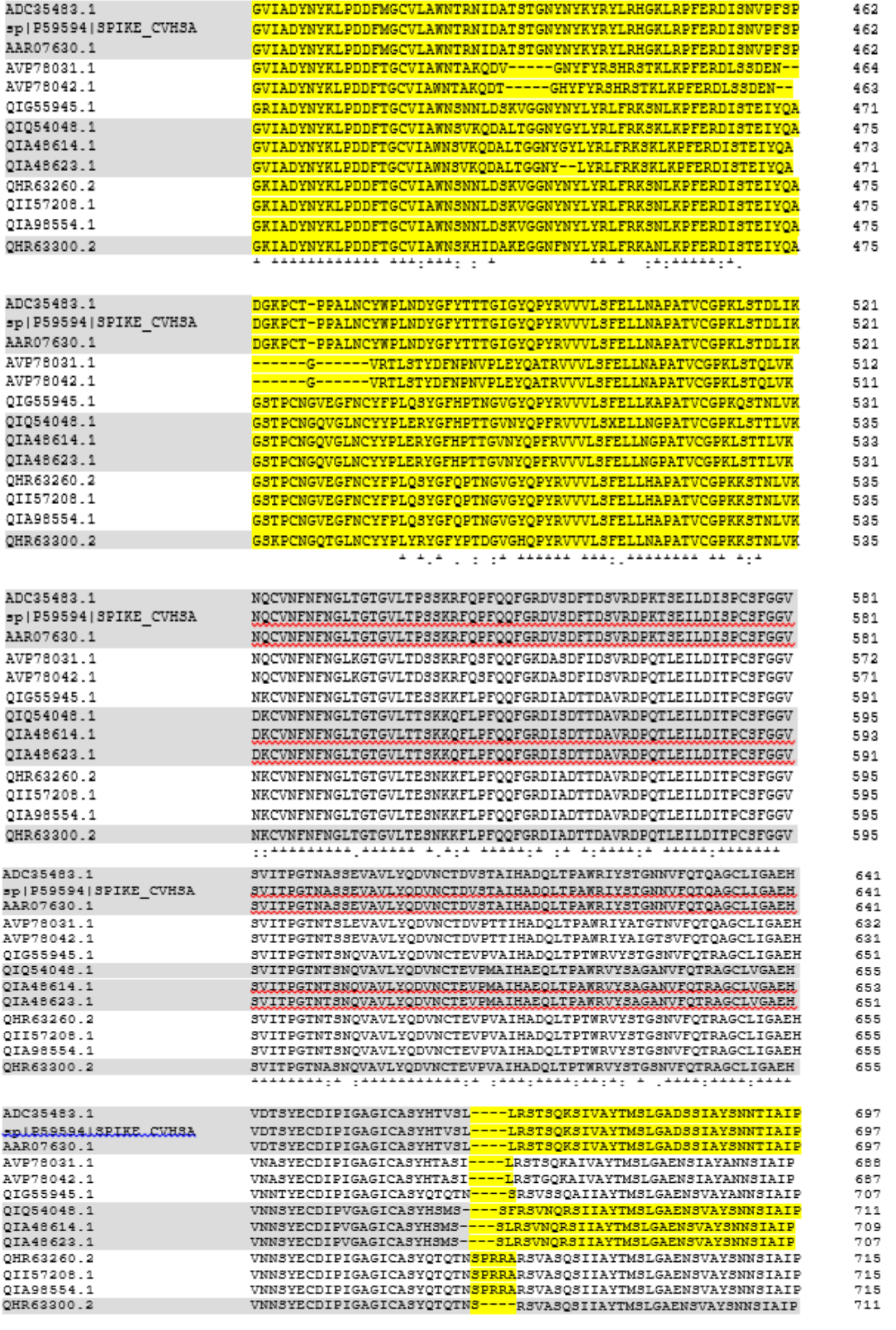
aligned genome sequences of 2019-nCoV, RaTG13, SARS-CoV and previously reported bat SARSr-CoVs, no evidence for recombination events was detected in the genome of 2019-nCoV (SARS-CoV-2)". Alternatively, it could be assumed that at some point BatCoV-RatG13 S gene, had been subject to heavy selective pressure, however, the functional constraints exceed the random drift or random chance.

Figure 3. Coronavirus spike glycoprotein multiple sequence alignment.

Spike glycoprotein multiple sequence alignment. Groups of sequence sample: (i) human SARS coronavirus (SARS-CoV) (1255 amoniacids length); (ii) Bat-SL-CoV (1245); (iii) Pangolin-CoVs (1265-1269); (iv) SARS-CoV-2 (1273); and (v) BatCoV-RaTG13 (1269). To better visualize the characteristics of each group, there were three representative sequences of each. The protein alignments were created by Clustal Omega (v.1.2.4) using default parameters (18). Strictly conserved amino acids are denoted by *, gaps are denoted by -. The characteristic deletion and insertions of the COVID-19 related sequences, the RBD, and the SARS-CoV-2 polybasic furin cleavage site are highlighted in yellow. The color of the bands (gray-white) are intended to highlight the characteristics of: SARS-CoV; Bat-SL-CoV and Pangolin-CoV (MP789); other Pangolin-CoVs; SARS-CoV-2; and BatCoV-RaTG13. The RBD position was based on (19). The figure only show up to the furin cleavage site. Up to the C-terminal end, most positions were strictly conserved. GenBank accession number of the spike glycoprotein sequences and the coronavirus are the following: ADC35483.1, SARS coronavirus HKU-39849; sp|P59594| (UNIPROT SPIKE_CVHSA); AAR07630.1, SARS coronavirus BJ302; AVP78031.1, Bat-SL-CoV ZC45, MG772933.1; AVP78042.1, Bat-SL-CoV ZXC21, MG772934.1; QIG55945.1, Pangolin-CoV, MP789,MT121216.1; QIQ54048.1, Pang-CoV,GX-P2V,MT072864.1; QIA48614.1, Pang-CoV,GX-P4L,MT040333.1; QIA48623.1, Pang-CoV,GX- P1E,MT040334.1; QHR63260.2, SARS-CoV-2, MN996528.1; QII57208.1 SARS-CoV-2, MT159709.2; QIA98554.1, SARS-CoV-2, MT066156.1; QHR63300.2, BatCoV-RaTG13, MN996532.1.







Cite this article: The dark side of COVID-19, Sci J Phar and Pharmaceu Sci. 2020; 2(1): 01-08.

Using a forensic genetic criterion, which is valid in a jury, in this case of the SARS-CoV-2 (due the presence of the polybasic furin cleavage site), there is a likelihood ratio (LR) that is "infinite". That is, the relationship between the probability (P1) that given the SARS-CoV- 2 genome sequence, this sequence belongs to SARS-CoV-2, obviously, it is 1; and the probability (P2) that given the same SARS-CoV-2 genome sequence, it belongs to another coronavirus of its taxonomic group, at the moment it is 0 (because the polybasic furin site has only been found in SARS-CoV-2). Then that LR is infinite (LR = P1/P2 = 1/0 = infinite). It is a probabilistic way that shows how difficult is to associate SARS-CoV-2 with lineage B of the beta-coronaviruses. In this sense, Z.-L. Shi and coworkers, 2020 (1), (not taking into account the furin site), also pointed out this doubt, literally: "Phylogenetic analysis of the full-length genome and the gene sequences of RdRp and spike (S) showed that—for all sequences—RaTG13 is the closest relative of 2019-nCoV (SARS-CoV-2) and they form a distinct lineage from other SARSr-CoVs".

Turning now to the the origin of the SARS-CoV-2, it must be taken into account that in the world of viruses the basic principles of biology are also fulfilled. From the Cell Theory of Rudolf Virchow (1858), "Omnis cellula ex cellula" (each cell derived from another pre- existing cell), it could be inferred as "each virus derives from a pre-existing virus". The Natural Selection of Charles Darwin (1859) of change mutations in the struggle for existence is fully applicable to the viruses. The principle of Theodosius Dobzhansky (1973) "Nothing in Biology Makes Sense Except in the Light of Evolution" (13) must be also applicable to viruses.

In science, and in biology, there have always been, there are and will be doubts and enigmas to be solved. Then, it is surprising that after almost a year of the pandemic, no more coronaviruses of BatCoV-RaTG13 species have been isolated. The bat (the natural virus reservoir most studied) strains could be isolated from fecal swab, and their complete genomes could be submitted to GenBank database for analysis by the scientific community. At least, this would serve to validate statistically the 96% genome identity with SARS-CoV-2. There is still 4%, which may hold the key to the presence of the polybasic furin site in the SARS-CoV-2 spike glycoprotein. As for SARS-CoV-2, there are thousands of complete genomes available in the database.

In biological evolution there are always doubts and "missing links", that based on the discovery of new fossil remains, could be cleared or resolved. However, doubt about the origin of SARS-CoV-2 is serious, because the doubt may include the possibility that its origin could be a laboratory construct or a purposefully manipulated virus. The right technology for this is available. Thus, for scientific and medical purposes, several synthetic constructs of SARS-CoV-2 genome exist in GenBank database (accession number MT108784.1, MT461669.1, MT461671.1, MT461670.1), or for therapeutic purposes, synthetic recombinant bat SARS-like coronavirus was created, and was infectious in cultured cells and in mice (14).

Finally, the availability of multiple BatCoV-RaTG13 complete genomes for comparative analysis would provide new perspective. Until this happens, the origin of SARS-CoV-2 is the dark side of COVID-19. There is reasonable hope, just for probability and because the natural "trial-error" evolutionary mechanism, that a human virus like SARS-CoV-2 will not appear again for a long, long time (nobody gets the lottery jackpot twice). Now we have to wait for an upcoming vaccine and/or specific therapeutic drugs available for COVID-19.

Acknowledgements

This work has not been awarded grants by any research-supporting institution

Author contributions

A.R. and E.O. performed project planning, coordination, execution, and facilitation. A.R. processed data collection and sequence and phylogenetic analysis. A.R. and E.O. prepared the manuscript.

Competing interest declaration

All authors declare that they have no conflicts of interest.

Cite this article: The dark side of COVID-19, Sci J Phar and Pharmaceu Sci. 2020; 2(1): 01-08.

References

1. Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao- Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao , Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Geng-Fu Xiao, Zheng-Li Shi. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273, 2020. PMID: 32015507. doi: 10.1038/s41586-020-2012-7.
2. Shauna Milne-Price, Kerri L Miazgowicz, Vincent J Munster. The emergence of the Middle East Respiratory Syndrome coronavirus. Pathog. Dis. 71:21-176, 2014. PMID: 24585737. doi: 10.1111/2049-632X.12166.
3 . Kristian G Andersen, Andrew Rambaut, W Ian Lipkin, Edward C Holmes, Robert F Garry. The proximal origin of SARS-CoV-2. Nat. Med. 26:450-452, 2020. PMID: 32284615. doi: 10.1038/s41591-020-0820-9.
4. Kimura, M. The neutral theory of molecular evolution. Cambridge University, Cambridge. UK (1983).
5. Yushun Wan, Jian Shang, Rachel Graham, Ralph S Baric, Fang Li. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade- Long Structural Studies of SARS Coronavirus. J. Virol. 94, e00127-20, 2020.PMID: 31996437. doi.org/10.1128/JVI.00127-20.
6. UNIPROT. P59594, Spike glycoprotein. Accessed October 09, 2020. https://www.uniprot.org/uniprot/P59594.
7. Markus Hoffmann, Hannah Kleine-Weber, Stefan Pöhlmann. Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Mol. Cell 78: 779–784, 2020. PMID: 32362314. doi: 10.1016/j.molcel.2020.04.022.
8 Dong-Sheng Chen, Yi-Quan Wu, Wei Zhang, San-Jie Jiang, Shan-Ze Chen. Horizontal gene transfer events reshape the global landscape of arm race between viruses and homo sapiens. Sci. Rep. 6:26934, 2016. PMID: 27270140. doi: 10.1038/srep26934.
9. Shahana S Malik, Syeda Azem-E-Zahra, Kyung Mo Kim, Gustavo Caetano-Anollés, Arshan Nasir. Do Viruses Exchange Genes across Superkingdoms of Life? Front. Microbiol. 8, 2110, 2017. PMID: 29163404. doi.org/10.3389/fmicb.2017.02110.
10. Shuai Xia, Qiaoshuai Lan, Shan Su, Xinling Wang, Wei Xu, Zezhong Liu, Yun
Zhu, Qian Wang, Lu Lu, Shibo Jiang. The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. Signal Transduct. Target Ther. 5:92, 2020. PMID: 32532959. doi.org/10.1038/s41392-020-0184-0.
11. Elisabeth Braun, Daniel Sauter. Furin-mediated protein processing in infectious diseases and cancer. Clin. Transl. Immunol. E1073, 2019. PMID: 31406574. doi.org/10.1002/cti2.1073.
12. Javier A Jaimes, Nicole M André, Joshua S Chappie, Jean K Millet, Gary R Whittaker. Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. J. Mol. Biol. 432. 3309-3325, 2020. PMID: 32320687. doi: 10.1016/j.jmb.2020.04.009.
13. Theodosius Dobzhansky. Nothing in Biology Makes Sense except in the Light of Evolution. The American Biology Teacher, 35: 125-129, 1973.
14. Michelle M Becker , Rachel L Graham, Eric F Donaldson, Barry Rockx, Amy C
Sims, Timothy Sheahan, Raymond J Pickles, Davide Corti, Robert E Johnston, Ralph S Baric, Mark R Denison. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. Proc. Natl. Acad. Sci. USA, 105: 19944–19949, 2008. PMID: 19036930. doi: 10.1073/pnas.0808116105
15. Muhamad Fahmi, Yukihiko Kubota, Masahiro Ito. Nonstructural proteins NS7b and NS8 are likely to be phylogenetically associated with evolution of 2019-nCoV. Infect. Genet. Evol. 81, 104272, 2020. PMID: 32142938. doi.org/10.1016/j.meegid.2020.104272.
16. Tao Zhang, Qunfu Wu, Zhigang Zhang. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr. Biol. 30:1346–1351.e2, 2020. PMID: 32315626. doi: 10.1016/j.cub.2020.03.022.
17. Fábio Madeira, Young Mi Park, Joon Lee, Nicola Buso, Tamer Gur, Nandana Madhusoodanan, Prasad Basutkar, Adrian R N Tivey, Simon C Potter, Robert D Finn, Rodrigo Lopez. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 47(W1):W636-W641, 2019. PMID: 30976793. doi: 10.1093/nar/gkz268.
18. Jun Lan, Jiwan Ge, Jinfang Yu, Sisi Shan, Huan Zhou, Shilong Fan, Qi Zhang, Xuanling Shi, Qisheng Wang, Linqi Zhang, Xinquan Wang. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature, 581:215-220, 2020. PMID: 32225176. doi:10.1038/s41586-020-2180-5.

Cite this article: The dark side of COVID-19, Sci J Phar and Pharmaceu Sci. 2020; 2(1): 01-08.