European Review for Medical and Pharmacological Sciences 2020; 24: 4558-4564

Bioinformatic analysis indicates that SARS-CoV-2 is unrelated to known artificial coronaviruses

T. DALLAVILLA[1], M. BERTELLI[1], A. MORRESI[2], V. BUSHATI[1], L. STUPPIA[3], T. BECCARI[4], P. CHIURAZZI[5,6], G. MARCEDDU[7]

[1]MAGI's Lab, San Felice del Benaco (Brescia), Italy
[2]Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, Perugia, Italy
[3]Dipartimento di Scienze Psicologiche, Umanistiche e del Territorio, Facoltà di Medicina e Scienze della Salute, Università "G. d'Annunzio", Chieti-Pescara, Italy
[4]Dipartimento di Scienze Farmaceutiche, Università di Perugia, Perugia, Italy
[5]Istituto di Medicina Genomica, Università Cattolica del Sacro Cuore, Roma, Italy
[6]Fondazione Policlinico Universitario "A.Gemelli" IRCCS, UOC Genetica Medica, Roma, Italy
[7]MAGI Euregio, Bolzano, Italy

Abstract. – OBJECTIVE: SARS-CoV-2 is responsible for the present coronavirus pandemic and some suggestions were made about its possible artificial origin. We, therefore, compared SARS-CoV-2 with such known viruses that were prepared in the laboratory and other relevant natural strains to estimate their genetic relatedness.

MATERIALS AND METHODS: BLAST and clustalW were used to identify and align viral sequences of SARS-CoV-2 to other animal coronaviruses (human, bat, mouse, pangolin) and related artificial constructs. Phylogenetics trees were then prepared using iTOL.

RESULTS: Our study supports the notion that known artificial coronaviruses, including the chimeric SL-SHC014-MA15 synthesized in 2015, differ too much from SARS-CoV-2 to hypothesize an artificial origin of the latter. On the contrary, our data support the natural origin of the COVID-19 virus, likely derived from bats, possibly transferred to pangolins, before spreading to man.

CONCLUSIONS: Speculations about the artificial origin of SARS-CoV-2 are most likely unfounded. On the contrary, when carefully handled, engineered organisms provide a unique opportunity to study biological systems in a controlled fashion. Biotechnology is a powerful tool to advance medical research and should not be abandoned because of irrational fears.

Key Words:
SARS-CoV-2, CoViD-19, SARS, Chimeric virus, SL-SHC014-MA15, Biotechnology.

Introduction

SARS-CoV-2, the name given to the coronavirus (CoV) that caused the recent worldwide pandemic, is the seventh such virus known to infect humans[1,2]. It belongs to the beta subgroup of the coronaviruses, which along with viruses of the alpha subgroup affect mammals, causing respiratory tract infections[3]. Although its case fatality rate is not very high compared to other viruses of the same group, like SARS-CoV, responsible for the 2002-2003 outbreak, SARS-CoV-2 has proven to be more infectious than e.g. the seasonal flu, allowing the virus to spread worldwide at a strikingly fast rate[4]. In fact, as of April 2, 2020 the WHO has reported 900.000 confirmed cases in 206 countries[5].

The SARS coronavirus appears to enter the cell by binding with its S (spike) protein to ACE2, a membrane-bound carboxypeptidase which, among other things, inactivates the octapepetide angiotensin II, counterbalancing the hypertensive and proinflammatory function of ACE. The structural basis of the binding between the viral S protein and ACE2 has been investigated in great detail[6-9] and the critical portion of the spike glycoprotein (known as Receptor-Binding Domain or RBD) has revealed 6 critical residues (Leu455, Phe486, Gln493, Ser494, Asn501, Tyr505) at the interface with ACE2[6]. Furthermore, a unique characteristic of the human SARS-CoV-2 coronavirus that may have improved its contagiousness is the insertion of 12-nucleotides resulting in 4 extra aminoacids (Pro-Arg-Arg-Ala) in position

Corresponding Author: Pietro Chiurazzi, MD, Ph. D; e-mail: pietro.chiurazzi@unicatt.it

681-684[10]; these 4 extra residues create a polybasic cleavage site unique to this human virus (not found in any other species nor in the 2002-2003 SARS-CoV virus) which may be involved in cleavage of the spike protein, facilitating entry in target cells. Sequence comparison at the nucleotide (as well as aminoacid) level is a useful technique to reconstruct the phylogenesis of viruses and this approach was first used by Li et al[10] and Paraskevis et al[11], allowing to trace the probable origin of the SARS-CoV-2 to the bat RaTG13 coronavirus. However, multispecies sequence alignment suggests that another host animal is involved in the transmission from bats to humans, possibly pangolin (*Manis javanica*) given the almost complete identity of the RBD between the human and pangolin virus (50/51 residues i.e., 98%) compared to the bat RaTG13 virus (41/51 residues i.e. 80%)[12,13]. Already 10 years ago Graham and Baric[14] nicely explained how coronaviruses are frequently characterized by host shifting events, whether they be animal-to-human (zoonosis), human-to-animal (reverse zoonosis), or animal-to-animal. Andersen et al[15] discussed in detail the possible scenarios of viral evolution and suggested that the insertion of the polybasic cleavage site may have occurred only in humans during a pre-epidemic period when human-to-human transmission went undetected because of a lower pathogenicity; studies of banked human samples might reveal if and when such a cryptic spread occurred.

Thus, although the evolution history of the SARS-CoV-2 virus is still being detailed, evidence points clearly to its natural origin from bats through other species (e.g. pangolin) to man. In spite of all the evidence, fuelled by fear and ignorance, conspiracy theories have emerged through the media from several political and scientific sources[16]. One such theory is based on a paper published on Nature in 2015[17], where a chimeric mouse/bat coronavirus was artificially assembled and used *in vitro* for studying if the S gene of the bat SL-SHC014 could enhance the infectiousness of the mouse MA-15 coronavirus. Another theory published on BioRivx (already withdrawn at the time of writing) suggested that the virus was a combination of a coronavirus and HIV-1 because of four insertions in the spike glycoprotein gene of SARS-CoV-2 that seemed identical to HIV-1. A rebuttal to these and other claims has been already provided[18-20]. Here we intend to provide further evidence for the natural origin of the SARS-CoV-2 virus and show that known artificial coronaviruses differ too much to be considered responsible for the COVID-19 pandemic.

Materials and Methods

Nucleotide sequences used in this study were selected from the literature and downloaded from GeneBank[21] (https://www.ncbi.nlm.nih.gov/genbank/) and GISAID[22] (https://www.gisaid.org/). The genomes of the patented coronaviruses US10130701B2 (https://patents.google.com/patent/US10130701B2/en) and US20070128224A1 (https://patents.google.com/patent/US20070128224A1/es) were manually searched and downloaded. US10130701B2, first patented in 2015 by the English Pirbright Institute, is an attenuated coronavirus comprising a variant replicase gene, which causes the virus to have reduced pathogenicity. US20070128224A1, first patented in 2004 upon request of the French CNRS, is a SARS-CoV derived from a sample collected in Hanoi (Vietnam). This strain underwent several modifications on different codons in the gene S and M of the SARS virus. A complete list of the sequences used in our study follows:

Patented coronaviruses:
- US10130701B2
- US20070128224A1
- Alpha coronaviruses
- 229E (AF304460.1)
- NL63 (MG428704.1)
- Beta coronaviruses lineage A
- OC43 (MN026164.1)
- HKU1 (NC_006577.2)
- Beta coronaviruses lineage B
- SARS CoV Urbani (AY278741.1)
- SARS-CoV MA-15 (DQ497008.1)
- Bat SARS-like-CoV RsSHC014 (KC881005.1)
- Bat BetaCoV/YN2018B (MK211376.1)
- Bat SARS-like-CoV Rs4231 (KY417146.1)
- RaTG13 Bat coronavirus (MN996532.1)
- Pangolin SARS-CoV P4L (EPI_ISL_410538)
- Pangolin SARS-CoV P5L (EPI_ISL_410540)
- Beta coronaviruses lineage C:
- MERS-CoV (MG987420.1)

The sequence for the chimeric SL-SHC014-MA15 virus[17] has been reconstructed *in silico* by substituting the coding sequence of the murine MA15 SARS-CoV S gene with the S gene from the bat SL-SHC014 coronavirus. BLASTn[23] with default parameters was used to compare the whole genome sequence of SARS-CoV-2 (NC_045512.2) with all the above viruses. The comparison was also performed using only the S gene of each virus. The coordinates of the S

gene (encoding the Spike glycoprotein of the viral envelope) were retrieved from GeneBank and GISAID (for pangolin viruses). Then, clustalW[24] was employed with default parameters for generating multiple alignment of the sequences listed above. Multiple alignments were then used to generate phylogenetic trees using iTOL[25].

Results

In order to further clarify the origin of SARS-CoV-2 we performed a series of comparison between its whole genome (NC_045512.2) and the genomic sequences of artificial coronaviruses developed in laboratory for relevant research purposes (e.g. vaccine or antibody production, improved understanding of viral pathogenesis). Two patented synthetic coronaviruses derived from samples of human SARS-CoV were taken into consideration (patent n. US20070128224A1 and US10130701B2), that were developed in 2004 and 2015 respectively. Both these strains were created for research purposes i.e. testing diagnostic reagents and vaccines development. We also reconstructed the chimeric SL-SHC014-MA15 virus, artificially synthesized in 2015 by inserting the S gene of bat RsSHC014 CoV in the mouse-attenuated MA-15 coronavirus of human origin[17]; this artificial virus proved very useful to understand which mutations can increase the pathogenicity of a coronavirus and allow its transfer from host animals to humans. Unfortunately, the SL-SHC014-MA15 virus was also implicated in a conspiracy theory widely credited on social media. In order to provide an adequate background, we also included in the comparison a number of human and animal coronaviruses as indicated in the previous section. Table I lists the one-to-one comparisons and sequence relative identity when the entire genome was aligned, while Table II reports the comparison only of the S gene, encoding the spike glycoprotein critical for viral infection.

In particular we compared to SARS-CoV-2: SARS-CoV Urbani (responsible for the 2002/2003 epidemic), MERS-CoV (responsible for the 2012 outbreak), NL63 and 229E (human alpha-CoV), HKU1 and OC43 (human beta-CoV lineage A), 2 pangolin CoV (GX/P4L, GX/P5L), 4 bat CoV (RaTG13, Rs4231, RsYN2018B, RsSHC014) and one murine (MA-15 SARS-CoV).

Table I confirms that the closest genome to SARS-CoV-2 is RaTG13 (96% identity), corroborating the hypothesis that the virus originated from bats. However, RaTG13 and human SARS-CoV-2 still differ at ~ 1000 bases, suggesting that an intermediate species was responsible for the zoonotic transfer to man[12,13]. The homology with RaTG13 is lower (93%) when only the S gene is compared (see Table II) and, as evidenced in Figure 1a of Andersen et al[15], 5 out of 6 critical residues in the ACE2 receptor binding domain

Table I. BLAST comparison with the whole genome sequence of SARS-CoV-2.

SARS-CoV-2 whole genome (NC_045512.2) compared to:	GeneBank/ GISAID	Origin	Max Score	Total Score	Query Cover	E value	% Identity
RaTG13 Bat coronavirus	MN996532.1	bat	48724	48724	99%	0	96.12%
SARS-CoV-2 pangolin GX/P5L	EPI_ISL_410540	pangolin	28301	31378	99%	0	85.98%
SARS-CoV-2 pangolin GX/P4L	EPI_ISL_410538	pangolin	28297	31369	99%	0	85.97%
BtRs-BetaCoV/YN2018B	MK211376.1	bat	15176	22618	91%	0	82.32%
Bat SARS-like Rs4231	KY417146.1	bat	15176	22534	91%	0	82.30%
SARS coronavirus Urbani	AY278741.1	human	15169	22535	88%	0	82.30%
US20070128224A1	—	**artificial**	**15169**	**22564**	**88%**	**0**	**82.30%**
SARS-CoV MA15	DQ497008.1	mouse	15151	22505	88%	0	82.28%
SARS-like SHC014-MA15 chimeric virus	—	**artificial**	**15104**	**22463**	**89%**	**0**	**82.24%**
Bat SARS-like SHC014	KC881005.1	bat	14938	22388	90%	0	82.07%
HKU1 (beta coronavirus lineage A)	NC_006577.2	human	496	496	5%	8.00E-140	72.49%
US10130701B2	-	**artificial**	**no significant correlation found**				

Natural strains with less than 70% of identity (% identity) were excluded from the results. Results show that bat RaTg13 is the virus closest to SARS-CoV-2, followed by the pangolin coronaviruses. All the artificial constructs, including the chimeric virus reported in 2015[17], as well as the bat RsSHC014 and the SARS-CoV MA-15 (shaded in gray) used for its creation, show a lower percentage of identity.

Table II. BLAST comparison with the S gene sequence of SARS-CoV-2.

SARS-CoV-2 gene S (from NC_045512.2) compared to:	GeneBank/ GISAID	Origin	Max Score	Total Score	Query Cover	E value	% Identity
RaTG13 Bat coronavirus	MN996532.1	bat	5541	5541	100.00%	0	92.89%
SARS-CoV-2 pangolin GX/P4L	EPI_ISL_410538	pangolin	3568	3568	100.00%	0	83.62%
SARS-CoV-2 pangolin GX/P5L	EPI_ISL_410540	pangolin	3563	3563	100.00%	0	83.59%
US20070128224A1	—	**artificial**	**1823**	**1823**	**74.00%**	**0**	**78.41%**
SARS coronavirus Urbani	AY278741.1	human	1823	1823	74.00%	0	78.41%
SARS-CoV MA15	DQ497008.1	mouse	1818	1818	74.00%	0	78.38%
BtRs-BetaCoV/YN2018B	MK211376.1	bat	1853	1853	79.00%	0	77.92%
Bat SARS-like SHC014	KC881005.1	bat	1831	1831	79.00%	0	77.80%
Bat SARS-like Rs4231	KY417146.1	bat	1808	1808	79.00%	0	77.68%
US10130701B2	—	**artificial**	**no significant correlation found**				

Natural strains with less than 70% of identity (% identity) were excluded from the results. Again, we find that RaTG13 is the closest to SARS-CoV-2 followed by the pangolin CoV. None of the SARS like viruses, included those from the work of Menachery et al[17] (shaded in gray) reaches a percentage of identity compatible with the hypothesis a non-natural origin of SARS-CoV-2.

(RBD) are different between SARS-CoV-2 and RaTG13. The second most similar genome is that of pangolins (86% identity), while all the other bat genomes and the mouse MA-15 virus have a homology of ~82% like that of SARS-CoV Urbani, with which they are related. Table I also shows that the US0070128224A1 patented virus, derived from human SARS-CoV, and the chimeric SL-SHC014-MA15 have ~82% identity i.e., between 5000 and 6000 nucleotides different from SARS-CoV-2: this indicates that none of these viruses is even a precursor of the virus causing COVID-19. Finally, the similarity with the other patented virus US10130701B2 is so low that BLAST finds no significant correlation.

The results of Table II, focusing only on the S gene, are similar to those of Table I, with bat RaTG13 CoV as the closest neighbour of SARS-CoV-2, with homology levels dropping 3-4%.

Then, after multiple sequence alignment with clustalW, we used iTOL to generate unrooted dendrograms graphically showing the phylogenetic distances between all viruses included in our work. Figure 1 shows these tree charts obtained by aligning the entire viral genome (a) or just the S gene of the virus (b). Relationship among the different viral sequences is similar in both charts and their relative similarity is inversely proportional to their distance; for example, the closeness between SARS-CoV-2 and RaTG13 is even more marked than just looking at the Tables. Furthermore, a clear clustering appears with a split between alpha and beta CoV and between the various lineages (A, B, C) of beta coronaviruses. The pangolin coronaviruses are then second closest to SARS-CoV-2, supporting the notion that this species may be involved in the natural evolution of the virus and the zoonotic transfer to human[13,15]. As far as the artificial constructs are concerned, US10130701B2 is farthest and clusters closer to beta CoV lineage A, while US20070128224A1 and the chimeric SHC014-MA15 are very close to the SARS-CoV Urbani (from which they derive) but enough distant from SARS-CoV-2 (approx. 5000-6000 different nucleotides) to exclude any possible involvement of these viruses in the development of SARS-CoV-2[19].

Discussion

Coronaviruses are enveloped RNA viruses responsible for infection and disease in many avian and mammal species. They contain the largest single-stranded, positive-sense RNA genomes currently known, ranging in size from 27 to nearly 32 kb in length. SARS-CoV-2 with 29903 nucleotides, encodes 10 different polypeptides, one of which corresponds to the S or "Spike" glycoprotein. The S protein is extremely important for the entry of the virus into target cells, including (but not limited to) type II pneumocytes. As discussed by Wan et al[6] the receptor binding domain (RBD) of the S protein is the critical determinant for species specificity and new variants acquired in the host species (or during a pre-epidemic phase of human-to-human transmission) may actually trigger the epidemic phase[10,15].

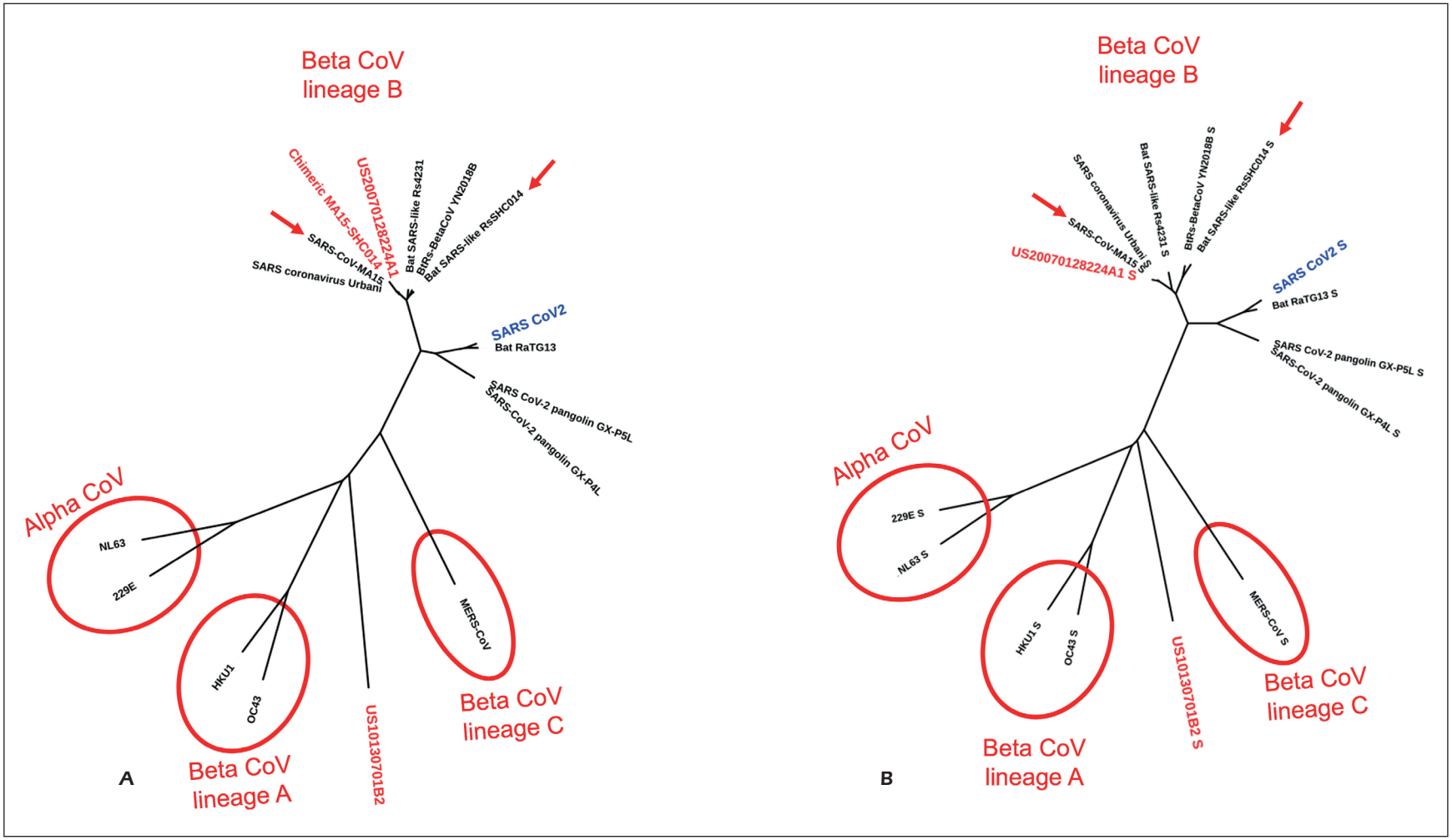

Figure 1. Unrooted trees for **A**) whole genomes and **B**) the S gene (encoding the spike glycoprotein) of the different coronaviruses analyzed in this study. Both trees show the close similarity of human SARS-CoV-2 and bat RaTG13 coronavirus. SARS-CoV Urbani (responsible for the SARS 2002-2003 outbreak), other bat strains and the artificial US20070128224A1 are more distantly related. Alpha CoV (229E, NL63), beta CoV lineage A (HKU1, OC43), MERS-CoV and the artificial US10130701B2 are even more unrelated to SARS-CoV-2. Please, note that the chimeric virus reported in 2015[17] and its components (murine SARS-CoV-MA15 and bat SARS-like SHC014, indicated by red arrows) cluster closer to SARS-CoV Urbani that circulated in 2002-2003 than to the present SARS-CoV-2 virus.

In the case of SARS-CoV-2 this "jump" to high pathogenicity may depend on the acquisition of the 4 extra amino acids in position 681-684 that introduce a polybasic cleavage site, found only in humans. Our results (BLAST alignment and phylogenetic tree reconstruction) are compatible with a natural evolutionary history of the human coronavirus causing COVID-19 pandemic; this is indicated by the high level of homology with the bat RaTG13 CoV and recently identified pangolin CoVs[13]. Dendrograms summarizing the reciprocal similarities among all investigated viral genomes are also compatible with the subspecies organization of Coronaviruses in alpha, beta, gamma (including the 3 different lineages of beta CoVs) and clearly show that artificial viral constructs, including the chimeric SHC014-MA15, are very different from the presently circulating SARS-CoV-2, excluding that this virus had been genetically engineered. Our results and other already quoted works[12,18-20] suggest that there is no plausible scientific evidence behind the hypothesis of a laboratory engineered virus. Studies of Menachery et al[17] and in general all research work involving genetically modified viruses should not be considered a threat, actually genetic engineering allows to greatly decrease research time required to develop an effective response to a disease. Moreover, all the genetic engineering studies need to follow very strict bio-security protocols, which greatly minimize hazards. International collaboration among researchers of different countries is of paramount importance in order to rapidly share all available information on emerging infectious diseases; a positive example is offered by the Global Initiative on Sharing All Influenza Data (GISAID)[22] that has been providing timely genetic information on the SARS-CoV-2 virus. Furthermore, molecular genetic techniques are extremely well suited to continuous population monitoring in order to reliably detect not only new human infectious diseases but also zoonoses currently affecting animals with which we have regular contact either dead (food) or alive (farming, breeding, company). Finally, many human infectious diseases originate from wildlife mammals, bats being a common host that is included in the dietary habits of many countries in Asia and Africa[26]. Therefore, more attention should be paid to the commerce of live animals in marketplaces to avoid animal-to-human and human-to-animal viral transmission. Strict rules should be set regarding the commerce of wild species when knowledge of their virome is poor and genetic surveillance should be implemented with regular sampling of animal food for the presence of potentially harmful viruses.

Conclusions

Our results integrate previous studies on pathogenic coronaviruses and clearly suggest that the development of SARS-CoV-2 followed natural selection and the hypothesis of its artificial origin in a laboratory is unfounded. We included in our analysis some of the available synthetic forms of coronaviruses and specifically excluded that the chimeric SHC014-MA15 virus[17] may be related to the SARS-CoV-2 underlying the present pandemic. On the other hand, the construction that chimeric virus[17] in 2015 resulted in a relevant advancement of our understanding of the pathogenic potential of a SARS-like virus potentially affecting humans. Biotechnology actually provides tools for rapidly elucidating disease mechanisms as well as for producing (in a faster and safer way) proteins and peptides "on demand" for antibody or vaccine production without handling active viral particles. Evidently, as with every powerful technique, careful planning and adequate protective measures must be in place to prevent accidental releases; however, irrational fears against biotechnology would only preclude access to better care and faster countermeasures to health threats like the SARS-CoV-2.

Acknowledgments

The Authors wish to thank Prof. Lucio Luzzatto and Prof. Michele Samaja for their critical reading of the manuscript and their precious suggestions.

Conflict of Interests

The Authors declare that they have no conflict of interests.

References

1) Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W; China Novel Coronavirus Investigating and Research Team. A Novel Coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020; 382: 727-733.

2) Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020; 579: 270-273.
3) Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF. Epidemiology, genetic recombination, and pathogenesis of Coronaviruses. Trends Microbiol 2016; 24: 490-502.
4) Callaway E, Cyranoski D, Mallapaty S, Stoye E, Tollefson J. The coronavirus pandemic in five powerful charts. Nature 2020; 579: 482-483. https://www.nature.com/articles/d41586-020-00758-2
5) Coronavirus diseases situation dashboard. [cited 2020 Mar 27]. Available at https://experience.arcgis.com/experience/685d0ace521648f8a5beeeee1b9125cd
6) 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: e00127-20.
7) Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 2020 Mar 6. pii: S0092-8674(20)30262-2. doi: 10.1016/j.cell.2020.02.058. [Epub ahead of print].
8) Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 2020; 367: 1260-1263.
9) Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol 2020; 5: 562-569.
10) Li X, Zai J, Zhao Q, Nie Q, Li Y, Foley BT, Chaillon A. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J Med Virol 2020 Feb 27. doi: 10.1002/jmv.25731. [Epub ahead of print]
11) Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. Infect Genet Evol 2020; 79: 104212.
12) Zhang C, Zheng W, Huang X, Bell EW, Zhou X, Zhang Y. Protein structure and sequence reanalysis of 2019-nCoV genome refutes snakes as its intermediate host and the unique similarity between its spike protein insertions and HIV-1. J Proteome Res 2020; 19: 1351-1360.
13) Lam TT, Shum MH, Zhu HC, Tong YG, Ni XB, Liao YS, Wei W, Cheung WY, Li WJ, Li LF, Leung GM, Holmes EC, Hu YL, Guan Y. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature 2020 Mar 26. doi:10.1038/s41586-020-2169-0. [Epub ahead of print]
14) Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J Virol 2010; 84: 3134-3146.
15) Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry, RF. The proximal origin of SARS-CoV-2. Nat Med 2020; 26: 450-452.
16) Pauls K, Yates J. Online claims that Chinese scientists stole coronavirus from Winnipeg lab have 'no factual basis'. CBC [Internet]. 2020 Jan 27 [cited 2020 Mar 27]. Available at https://www.cbc.ca/news/canada/manitoba/china-coronavirus-online-chatter-conspiracy-1.5442376
17) Menachery VD, Yount BL Jr, Debbink K, Agnihothram S, Gralinski LE, Plante JA, Graham RL, Scobey T, Ge XY, Donaldson EF, Randell SH, Lanzavecchia A, Marasco WA, Shi ZL, Baric RS. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nat Med 2015; 21: 1508-1513.
18) Xiao C, Li X, Liu S, Sang Y, Gao SJ, Gao F. HIV-1 did not contribute to the 2019-nCoV genome. Emerg Microbes Infect 2020; 9: 378-381.
19) Liu S, Saif LJ, Weiss SR, Su L. No credible evidence supporting claims of the laboratory engineering of SARS-CoV-2. Emerg Microbes Infect 2020; 9: 505-507.
20) Hao P, Zhong W, Song S, Fan S, Li X. Is SARS-CoV-2 originated from laboratory? A rebuttal to the claim of formation via laboratory recombination. Emerg Microbes Infect 2020; 9: 545-547.
21) Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2016 Jan 4;44(D1): D67-72.
22) Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall 2017; 1: 33-46.
23) Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol 2000; 7:203-214.
24) Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22: 4673-4680.
25) Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 2019; 47: W256-W259.
26) Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. Nature 2017; 546: 646-650.