Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Biosafety and Health 3 (2021) 124–127

Contents lists available at ScienceDirect

Biosafety and Health

journal homepage: www.elsevier.com/locate/bsheal

Genetic diversity and genomic epidemiology of SARS-CoV-2 in Morocco

Bouabid Badaoui *, Khalid Sadki, Chouhra Talbi, Driss Salah, Lina Tazi

Faculty of Sciences, Mohamed V University in Rabat, Morocco

ARTICLE INFO

Article history:
Received 23 June 2020
Received in revised form 18 January 2021
Accepted 29 January 2021
Available online 3 February 2021

Keywords:
SARS-CoV-2
Genetic diversity
Genomic epidemiology
Morocco

ABSTRACT

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), declared as a pandemic due to its rapid spread worldwide. In this study, we investigate the genetic diversity and genomic epidemiology of SARS-CoV-2, using 22 virus genome sequences reported by three different laboratories in Morocco till June 7,2020, as well as 40,366 virus genomes from all around the world. The SARS-CoV-2 genomes from Moroccan patients revealed 62 mutations, of which 30 were mis-sense mutations. The mutations Spike_D614G and NSP12_P323L were present in all the 22 analyzed sequences, followed by N_G204R and N_R203K, which occurred in 9 among the 22 sequences. The mutations NSP10_R134S, NSP15_D335N, NSP16_I169L, NSP3_L431H, NSP3_P1292L and Spike_V6F occurred once in Moroccan sequences, with no record in other sequences worldwide. Phylogenetic analyses revealed that Moroccan SARS-CoV-2 genomes included 9 viruses belonging to Clade 20A, 9 to Clade 20B and 2 to Clade 20C, suggesting that the epidemic spread in Morocco did not display a predominant SARS-CoV-2 route. Therefore, multiple and unrelated introductions of SARS-CoV-2 into Morocco through different routes have occurred, giving rise to the diversity of virus genomes in the country. Further, in all probability, the SARS-CoV-2 circulated in a cryptic way in Morocco, starting from January 15, 2020 before the first case was officially discovered on March 2, 2020.

1. Introduction

An outbreak of respiratory illness, named coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was reported, for the first time, in Wuhan (Hubei Province, China) in early December 2019 and subsequently many cases were reported in other countries worldwide. On 30 January, 2020, WHO declared that the corona-virus disease constituted a Public Health Emergency of International Concern.

This pandemic has infected more than 7,725,457 people globally and caused more than 427,683 deaths as of June 7, 2020 (https://www.worldometers.info/coronavirus/). This disease was confirmed to have spread to Morocco on 2 March, 2020, when the first COVID-19 case was confirmed. Until June 7, 2020, Morocco reported 8,692 confirmed cases and 212 deaths. The most important feature of this disease is expressed by a high-level of inflammatory response, including pro-inflammatory cytokines in an especially severe form which cause pneumonia and severe acute respiratory syndrome [1]. SARS-CoV-2 genome has a size of 29.8–29.9 kb, harboring a long 5′ end that contains orf1ab. The 3′ end codes for the structural proteins: small envelope (E) protein, matrix (M) protein, spike (S) protein and nucleocapsid (N) protein. Further, the SARS-CoV-2 genome contains six accessory proteins, encoded by ORF3a, ORF6, ORF7a, ORF7b, and ORF8 genes [2]. Compared to SARS-CoV, SARS-CoV-2 has high transmission and less pathogenicity [3] due to hitherto unknown reasons.

Virus evolution in nature arises through manifold mechanisms, of which nucleotide substitution is of utmost importance [4]. Genomic epidemiology of emerging viruses is of major relevance for capturing virus evolution and spread [5,6]. This approach has been proved to be very efficient during the Ebola virus epidemic in West Africa [7] and the Zika virus spread in Brazil [8].

To investigate the mutations underlying the evolution of SARS-CoV-2 in Morocco, 40,390 genome sequences of SARS-CoV-2, of which 22 were from Moroccan patients, were collected from GISAID (https://www.gisaid.org/) to study the diversity and genomic epidemiology of SARS-CoV-2 in Morocco.

2. Materials and methods

To assess the genetic variation of SARS-CoV-2 in Morocco, a total of 40,390 complete genomes of SARS-CoV-2 and their corresponding metadata were retrieved from GISAID database [9] and analyzed using the Nextstrain bioinformatics pipeline [10].

The platform Augur was used to perform a multi-alignment using all the genomes through MAFFT [11] and 'Wuhan-Hu-1/2019', and 'Wuhan/

* Corresponding author at: Faculty of Sciences, Mohamed V University in Rabat, Morocco.
E-mail address: bouabidbadaoui@gmail.com (Bouabid Badaoui).

http://dx.doi.org/10.1016/j.bsheal.2021.01.003

HIGHLIGHTS

Scientific question

This study investigated the diversity and genomic epidemiology of SARS-CoV-2 in Morocco.

Evidence before this study

To apply such approach for investigating the evolution of SARS-CoV-2 in Morocco, we collected from GISAID (https://www.gisaid.org/) 40,390 genome sequences of SARS-CoV-2, of which 22 were from Moroccan patients.

New findings

The genetic analysis of the SARS-CoV-2 genomes from Moroccan patients revealed some new mutations with no aforementioned record in other sequences worldwide. Genomic Epidemiology analyses revealed that the COVID-19 spread occurred through multiple and unrelated introductions of COVID-19 into Morocco via different routes.

Significance of the study

Deep genomic surveillance of SARS-CoV-2 reveals the circulating lineages of the virus, their changes over time and, in case of extensive sampling, could be a good way to evaluate how the interventions are affecting virus evolution.

WH01/2019' as reference genomes. The phylogeny was built by maximum likelihood using IQTREE [12]. For comparative analysis of SARS-CoV-2 genome sequences, we used the following protocol: i) Collection of SARS-CoV-2 sequences and the corresponding metadata from the GISAID database. ii) Filtering the SARS-CoV-2 sequences to exclude inaccurate sequences based on missing bases and sequence length and to set a fixed number of samples per group according to their similarities. iii) Performing a multi-sequence alignment via MAFFT. vi) Inferring a phylogenetic tree from the multi-sequence alignment, getting a time-resolved tree through TreeTime and inferring ancestral traits sequences, and v) Identifying the mutations.

3. Results

3.1. Genetic diversity and evolution of SARS-CoV-2 in Morocco

In this study, the analysis of 22 genomes from Moroccan patients revealed 62 mutations (Supplementary Table 1) of which 30 were missense mutations (Supplementary Table 2). Among these, Spike_D614G and NSP12_P323L were present in all the 22 analyzed sequences, followed by N_G204R and N_R203K occurring in 9 of the 22 sequences (Table 1).

Until June 7, 2020, the mutations Spike_D614G and NSP12_P323L had occurred 23,612 and 23,543 times respectively throughout 75 countries (https://www.gisaid.org/). Further, the mutations N R203K and N G204R have occurred, till now, 8,744 and 8,715 times respectively in 65 countries (https://www.gisaid.org/).

The mutation Spike M1237I, found just once in our sequences, had occurred earlier, 11 times worldwide in six countries.

The mutations NSP12_M196I, NSP3_A1819V, M_L13F, NSP14_D324A, NSP14_T75I and NSP5_V125I were found once in our sequences and were reported in only one sequence in the GISAID database. Therefore, NSP12_M196I, NSP3_A1819V, M_L13F, NSP14_D324A, NSP14_T75I and NSP5_V125I occurred once in India (hCoV-19/India/GBRC46/2020), Poland (hCoV-19/Poland/PL_P15/2020), Switzerland/100144/2020; United Kingdom / England EPI_ISL_423899, Australia, EPI_ISL_426742 and South America EPI_ISL_445334.

Table 1
Most frequent non-synonymous mutations in SARS-CoV-2 isolates from Moroccan patients.

SNP mutation	Protein mutation	Frequency
14408C > T	RNA pol (nsp12, P323L)	22
23403A > G	spike glycoprotein (S protein), D614G	22
28881G > A; 28882G > A	nucleocapsid phosphoprotein (R203K)	9
28883G > C	nucleocapsid phosphoprotein (G204R)	9

The mutations NSP10_R134S, NSP15_D335N, NSP16_I169L, NSP3_L431H, NSP3_P1292L and Spike_V6F were highlighted once in Moroccan sequences, with no record in other sequences around the world.

3.2. Genomic epidemiology of SARS-CoV-2 in Morocco

We now report the putative history of SARS-CoV-2 transmission in Morocco, as revealed by genomic epidemiology (Fig. 1 and Fig. S1). The virus entered and started circulating in Morocco, around the beginning of February, more likely from Belgium, Spain and France (Clade 20C was the first one to be introduced into Morocco); around March 4, 2020, new infected cases came from Belgium to Morocco (Clade 20A), around March 12, 2020, other infected persons entered Morocco from France and passing through Spain (Clade 20A). Three days later (March 14, 2020), infected cases reached Morocco from Switzerland (Clade A). Around March 17, 2020, other SARS-CoV-2 strains entered Morocco from USA (Clade 20A and 20B) and Germany (Clade 20B). Around March 22, 2020, travel- associated cases landed in Morocco from China (Clade A). Other infected arrivals were registered between the 16 and 19 April, 2020 from USA and Vietnam, more likely through sea trades that continued during the lockdown.

The virus genomes from Moroccan patients held between 4 and 16 mutations relative to the common ancestor [Wuhan-Hu-1/2019′, 'Wuhan/WH01/2019'] (Fig. 2). This pattern got a molecular evolution rate of 25.331 substitutions per year, consistent with what was reported before [13].

Phylogenetic analysis (Fig. 1) shows that the SARS-CoV-2 genomes from Moroccan patients are dispersed across the evolutionary tree of SARS-CoV-2, estimated from 40,390 genomes available on GISAID as of June 7, 2020. These included 9 viruses from Clade 20A, 9 from Clade 20B and 2 from Clade 20C (Fig. 3). This reveals that the epidemic spread in Morocco did not show a predominant SARS-CoV-2 lineage. However, it is more likely that the virus circulated covertly around the beginning of February, before the official discovery of the first case on March 2.

4. Discussion

Among the missense mutations, Spike_D614G, NSP12_P323L, N_R203K and N_G204R occurred with high frequency worldwide. It is very likely that this contributes to increase the SARS-CoV-2 transmissibility. Throughout its evolution within the host, the virus seeks to proliferate efficiently, while concurrently circumventing host morbidity to set out a maximum transmission [14]. This is in concordance with the concomitant reduced pathogenicity that accompanies its transmission increase. The Spike_D614G mutation manifested a major effect on the efficiency of the virus to infect hosts [15] and showed a high ability to hinder the immune systems of hosts that already dealt with version of SARS-CoV-2 without the Spike_D614G mutation [16]. This aspect should be emphasized for future vaccine researches. Though the mutation NSP12_P323L aroused the substitution of proline that plays a prominent role in protein folding and aggregation, neither increased the SARS-CoV-2's infectivity or its fitness regarding natural selection [17]. This might be because the change from proline to leucine amino-acid (P323L) did not change the protein function, as both amino-acids pertain to the non-polar aliphatic R groups.

Other missense mutations occurred either with small frequency like Spike_M1237I that was reported 11 times all over the world or those reported only once worldwide, like NSP12_M196I, NSP3_A1819V,

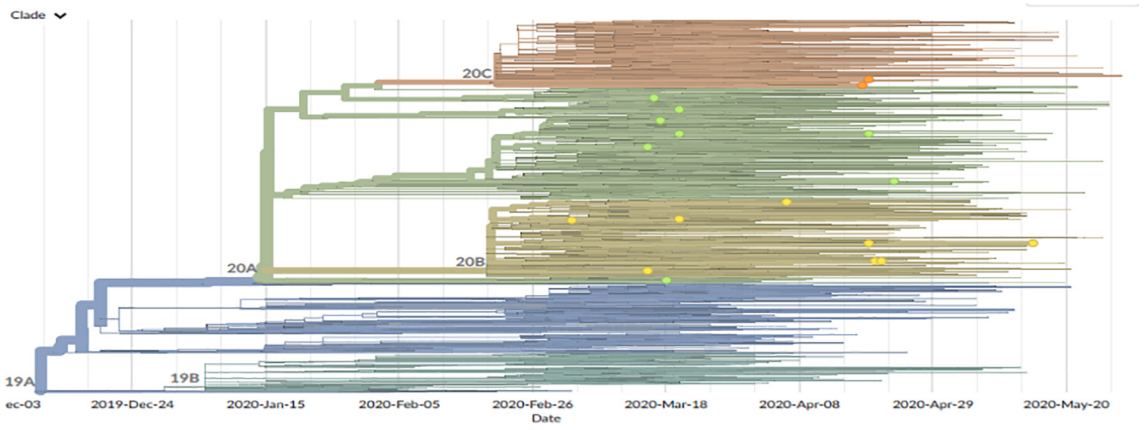

Fig. 1. Phylogeny of 40,390 SARS-CoV-2 genomes collected from GISAID database. The viruses collected from Morocco are highlighted, in colorful dots, according to the clades to which they belong. Clustering of related viruses indicates community transmission after an introduction event. Morocco witnessed three major transmission events.

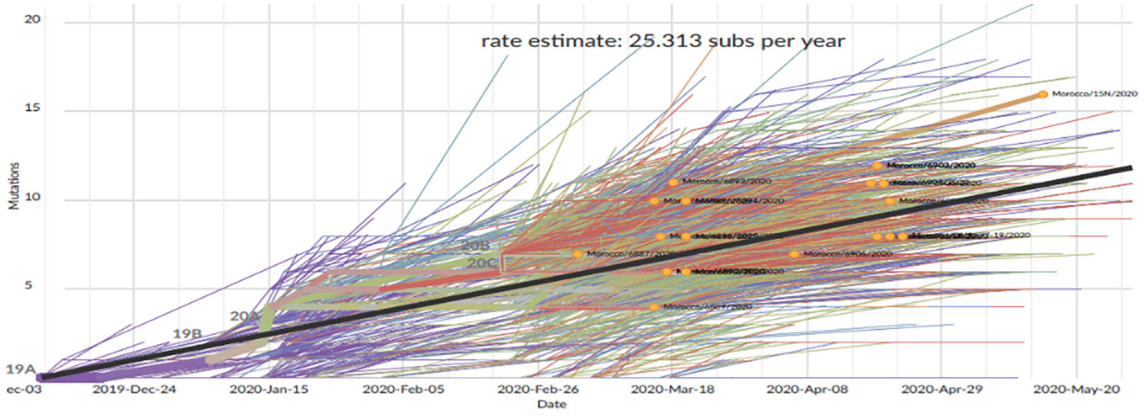

Fig. 2. Clock Diagram showing the evolution of mutations number in SARS-CoV-2 strains. The viruses collected in Morocco, colored in orange, got between 4 and 16 mutations compared to the reference sequence.

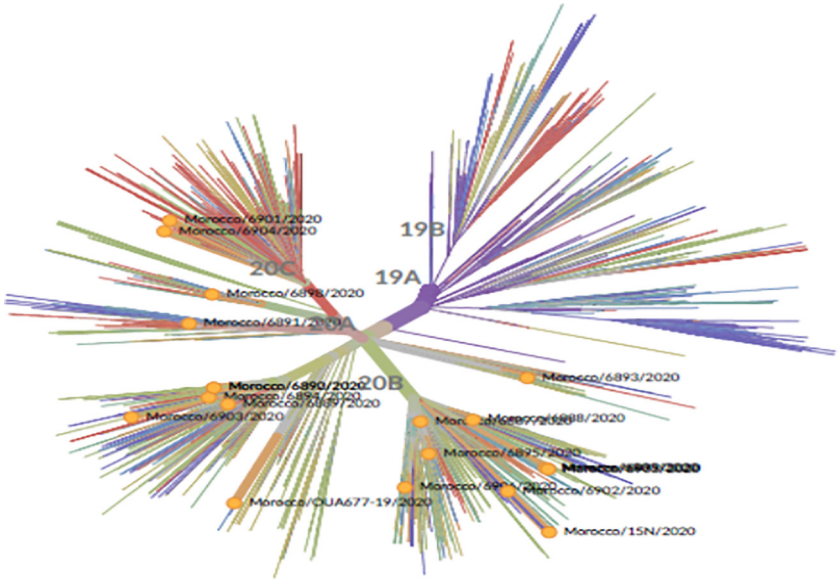

Fig. 3. Radial tree for the Phylogeny of 40,390 SARS-CoV-2 genomes collected from GISAID database. The viruses collected from Morocco, highlighted, in orange dots, belong to three clades 20A, 20B and 20C.

M_L13F, NSP14_D324A, NSP14_T75I and NSP5_V125I. These mutations seem to be rare and may represent the virus adaptation in the face of specific genetic backgrounds of the host, climate conditions or other unknown factors. Concerning the mutation Spike_M1237I, it's noteworthy that the spike surface glycoprotein is crucial for the virus binding to receptors on the host cell [18]. This glycoprotein is also the major target of neutralizing antibodies [19]. Hence, mutations in the spike surface glycoprotein could provoke change in the antigenicity of SARS-CoV-2.

The mutations NSP10_R134S, NSP15_D335N, NSP16_I169L, NSP3_L431H, NSP3_P1292L and Spike_V6F that occurred in the Moroccan sequences with no record in other sequences worldwide, should get careful attention and should be investigated to figure out their potential effects on the SARS-CoV-2 virulence, as well as their association with immunological and clinical symptoms. Special focus should be afforded to the mutation Spike_V6F, as the structural protein spike has been proven to be essential for the virus' ability to infect the hosts and was used as an important target for vaccine development [20].

Genomic epidemiology, using Nextstrain, applied to SARS-CoV-2 transmission in Morocco, revealed many aspects of the epidemic already known to the authorities, like the introduction of SARS-CoV-2 strains into Morocco from Belgium, Spain and France at the beginning of the epidemic. However, this analysis highlighted many other so far unknown events, like the arrival of strains from USA and Vietnam after the lockdown, possibly through sea trades.

SARS-CoV-2 genomes from Moroccan patients are dispersed across the evolutionary tree of SARS-CoV-2, with 9 viruses belonging to Clade 20A, 9 to Clade 20B and 2 to Clade 20C. This suggests that multiple and unrelated introductions of COVID-19 into Morocco have occurred through different routes, giving rise to the diversity of virus lineages reported in this study. This finding suggests that different SARS-CoV-2 strains, with different mutation patterns, coexist in Morocco. The contribution of each of those mutations needs to be investigated, in order to ascertain possible drug resistance and eventually different SARS-CoV-2 mortality rates related to these mutations.

It is tempting to speculate that the specific evolutionary profiles of the SARS-CoV-2 in Morocco might be the product of the interaction between its evolutionary routes before reaching the country and its adaptation to the Moroccan genetic background.

This genomic survey of SARS-CoV-2 in Morocco has at least two limitations. First, the sample size is relatively small, which might hinder a compelling interpretation of the epidemiological situation in the country. Albeit samples come from geographically diverse regions in Morocco, undersampling of virus genomes makes it hard to acquire a compelling picture of SARS-CoV-2 transmission. Deep genomic surveillance of SARS-CoV-2, with an extensive sampling, reveals the circulating lineages of the virus, their changes over time, as well as how the interventions are affecting virus evolution. However, the limited diagnostic and sequencing infrastructure in the country has seriously hindered the epidemic management and unquestionably, genomic and bioinformatics infrastructure has to be deployed to handle future outbreaks more effectively. It is paramount to learn the lessons from COVID-19 and develop the necessary tools for the next pandemic preparedness. Second, all the samples analyzed, except one, were obtained from public health laboratories and thus may not be representative of the general population.

5. Conclusions

The genetic analysis of the SARS-CoV-2 genomes from Moroccan patients revealed some new mutations, with no record in other sequences worldwide. These mutations should be investigated to ascertain their potential effects on the SARS-CoV-2 virulence and to evaluate their impacts on the immune response. The phylogenetic analyses revealed that the COVID-19 spread occurred through multiple and unrelated introductions of COVID-19 into Morocco through different routes. SARS-CoV-2 might have circulated in a covert way in Morocco around the beginning of February, before the discovery of the first case on March 2, 2020.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Author contributions

Bouabid Badaoui: Conceptualization, Formal Analysis, Methodology, Writing - Original Draft. **Khalid Sadki:** Writing - Review & Editing. **Chouhra Talbi:** Writing - Review & Editing. **Driss Salah:** Data Curation, Writing - Original Draft. **Lina Tazi:** Conceptualization, Writing - Review & Editing.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bsheal.2021.01.003.

References

[1] K.G. Andersen, A. Rambaut, W.I. Lipkin, et al., The proximal origin of SARS-CoV-2, Nat. Med. 26 (2020) 450–452, https://doi.org/10.1038/s41591-020-0820-9.
[2] R.A. Khailany, M. Safdar, M. Ozaslan, Genomic characterization of a novel SARS-CoV-2, Gene Rep. 19 (2020), 100682. https://doi.org/10.1016/j.genrep.2020.100682.
[3] Y. Fu, Y. Cheng, Y. Wu, Understanding SARS-CoV-2-mediated inflammatory responses: from mechanisms to potential therapeutic tools, Virol. Sin. 35 (2020) 266–271, https://doi.org/10.1007/s12250-020-00207-4.
[4] A.S. Lauring, R. Andino, Quasispecies theory and the behavior of RNA viruses, PLoS Pathog. 6 (2010), e1001005. https://doi.org/10.1371/journal.ppat.1001005.
[5] C. Fraser, C.A. Donnelly, S. Cauchemez, et al., Pandemic potential of a strain of influenza a (h1n1): early findings, Science. 324 (2009) 1557–1561, https://doi.org/10.1126/science.1176062.
[6] J.L. Gardy, N.J. Loman, Towards a genomics-informed, real-time, global pathogen surveillance system, Nat. Rev. Genet. 19 (2017) 9–20, https://doi.org/10.1038/nrg.2017.88.
[7] S. Baize, D. Pannetier, L. Oestereich, et al., Emergence of Zaire ebola virus disease in Guinea, N. Engl. J. Med. 371 (2014) 1418–1425, https://doi.org/10.1056/nejmoa1404505.
[8] N.R. Faria, J. Quick, I.M. Claro, et al., Establishment and cryptic transmission of zika virus in Brazil and the americas, Nature 546 (2017) 406–410, https://doi.org/10.1038/nature22401.
[9] Y. Shu, J. McCauley, GISAID: global initiative on sharing all influenza data from vision to reality, Euro. Surveill. 22 (2017) 30494, https://doi.org/10.2807/1560-7917.es.2017.22.13.30494.
[10] J. Hadfield, C. Megill, S.M. Bell, et al., Nextstrain: real-time tracking of pathogen evolution, Bioinformatics. 34 (2018) 4121–4123, https://doi.org/10.1093/bioinformatics/bty407.
[11] K. Katoh, MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform, Nucleic Acids Res. 30 (2002) 3059–3066, https://doi.org/10.1093/nar/gkf436.
[12] L.T. Nguyen, H.A. Schmidt, A. von Haeseler, et al., IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2014) 268–274, https://doi.org/10.1093/molbev/msu300.
[13] A. Rambaut, Phylogenetic analysis of nCoV-2019 genomes, University of Edinburgh, Edinburgh UK, 2020. https://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356 (accessed 10 June 2020).
[14] S. Alizon, A. Hurford, N. Mideo, et al., Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future, J. Evol. Biol. 22 (2008) 245–259, https://doi.org/10.1111/j.1420-9101.2008.01658.x.
[15] I.J.M. Júnior, R.C. Polveiro, G.M. Souza, et al., The global population of SARS-CoV-2 is composed of six major subtypes, Sci. Rep. 18289 (2020). https://doi.org/10.1038/s41598-020-74050-8.
[16] L. Zhang, C.B. Jackson, H. Mou, et al., SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity, Nat. Commun. 11 (2020), 6013, https://doi.org/10.1038/s41467-020-19808-4.
[17] A. Maitra, M.C. Sarkar, H. Raheja, et al., Mutations in SARS-CoV-2 viral RNA identified in Eastern India: possible implications for the ongoing outbreak in india and impact on viral structure and host susceptibility, J. Biosci. 45 (2020) 76, https://doi.org/10.1007/s12038-020-00046-1.
[18] T.S. Fung, D.X. Liu, Human coronavirus: host-pathogen interaction, Annu. Rev. Microbiol. 73 (2019) 529–557, https://doi.org/10.1146/annurev-micro-020518-115759.
[19] N. Chen, M. Zhou, X. Dong, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet 395 (2020) 507–513, https://doi.org/10.1016/s0140-6736(20)30211-7.
[20] S.F. Ahmed, A.A. Quadeer, M.R. McKay, Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies, Viruses 12 (2020) 3–254, https://doi.org/10.1101/2020.02.03.933226.