



***In silico* Molecular Characterization and Phylogenetic Analyses of SARS-CoV-2 in Mediterranean Basin**

Ghaleb. M. Adwan^{1*}

¹Department of Biology and Biotechnology, An-Najah National University, P.O. Box.7, Nablus, Palestine.

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/JAMB/2021/v21i430340

Editor(s):

- (1) Dr. Foluso O. Osunsanmi, University of Zululand, South Africa.
- (2) Dr. Muhsin Jamal, Abdul Wali Khan University, Pakistan.
- (3) Prof. Pongsak Rattanachaikunsopon, Ubon Ratchathani University, Thailand.

Reviewers:

- (1) Milene Oliveira de Souza, Centro Universitário AGES (UniAGES), Brazil.
- (2) Blanca Estela Chávez Sandoval, Universidad Autónoma Metropolitana (UAM), México.
- (3) Fabrício Vieira Cavalcante, Universidade de Brasília, Brazil.
- (4) Jesús G. Rodríguez Diego, Universidad Autónoma Metropolitana, Mexico.
- (5) Margeli Silva, Lisandro Alvarado University Centroccidental, Venezuela.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/67086>

Received 17 March 2021

Accepted 14 April 2021

Published 24 April 2021

Original Research Article

ABSTRACT

Background: The novel human coronavirus disease COVID-19 has become the fifth reported pandemic since the global spread of 1918 flu and counted as the first documented coronavirus worldwide spread in the history. The COVID-19 was initially considered as a respiratory disease, but SARS-CoV-2 can lead to cause other serious complications.

Purpose: This study aimed to conduct phylogenetic analyses of the whole genome of SARS-CoV-2 strains isolated from infected humans in Mediterranean basin countries, Orf1ab gene, S gene, M gene, N gene and Orf3a gene sequences. In addition, the products of Orf1ab, S, M and N genes were also phylogenetically analyzed. Changes that occurred on the S-gene product of these SARS-CoV-2 strains were also detected.

Materials and Methods: The whole genome of SARS-CoV-2 isolates, the genes and the gene products (Accessed July 20, 2020) were recovered in Mediterranean basin countries were retrieved from GenBank Database previously available in National Center for Biotechnology Information (NCBI) using BLAST (Basic Local Alignment Search Tool) system. Analyses of these sequences were carried out using computer program MEGA6 software.

*Corresponding author: E-mail: adwang@najah.edu;

Results: The Phylogenetic analyses showed that Bat coronavirus RaTG13 isolate is more closely related to SARS-CoV-2 isolates than Pangolin coronavirus isolates. The S gene product of this virus mediates entry into the host cell and has S1/S2 cleavage site containing multibasic amino acid sequence (PRRAR) which is not detected in other closely related coronaviruses. Many coronavirus strains that deposited in GenBank, showed that they have PRR sequence in Orf1ab gene product.

Conclusion: we conclude that part of multibasic S1/S2 motif acquired by recombination or insertion. Theoretically, any coronavirus strain acquired this sequence becomes highly pathogenic to humans. The dominant mutation (79.3%) at S gene product level was 614D→G. The impact of mutations detected in S gene product on virus transmission, diagnosis, pathogenicity and strategies of antiviral therapy, should be rapidly assessed in further studies.

Keywords: SARS-CoV-2; phylogenetic analyses; mediterranean basin countries; bat coronavirus RaTG13; S1/S2 cleavage site.

1. INTRODUCTION

The novel human coronavirus disease COVID-19 has become the fifth reported pandemic since the global spread of 1918 flu and counted as the first reported coronavirus worldwide spread in the history [1]. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a new formerly identified as the 2019 novel Coronavirus (2019-nCoV), is a recently emerging zoonotic pathogen which was initially documented in Wuhan China in December 2019. This virus causes the Coronavirus Disease 2019 (COVID-19) which involves severe respiratory diseases and other serious complications [2-4].

Although of SARS-CoV-2 virus is evolving, to date seven strains of human pathogenic coronaviruses are recognized. These include OC43, NL63, 229E, HKU1, severe acute respiratory syndrome (SARS-CoV), Middle East respiratory syndrome (MERS-CoV), and 2019-nCoV, responsible for the respiratory tract infections and other system symptoms. Other SARS related viruses were reported to be responsible for past epidemics including The SARS-CoV epidemic in China in 2002-2003, and the MERS-CoV epidemic in Kingdom of Saudi Arabia and other countries in the region in 2012-2013 [1,5-6].

SARS-CoV-2 belongs to the subgenus *Sarbecovirus*. Together with other subgenera such as *Embecovirus*, *Hibecovirus*, *Merbecovirus*, and *Nobecovirus*, that belong the genus *Betacoronavirus* (β CoV) (subfamily *Coronavirinae*, family *Coronaviridae*, suborder *Cornidovirineae* and order *Nidovirales*) [2]. In addition to the genus β CoV, there are other 3 genera α CoV, δ CoV and γ CoV.

Phylogenetic analyses have demonstrated that rodents and bats are the sources of most α CoVs and β CoVs, while avian species are considered

the major source of most γ CoVs and δ CoVs. The CoVs have repeatedly crossed species barriers, and some have emerged to become as a serious human pathogen [3].

The SARS-CoV-2 is a spherical enveloped particle, approximately 60-140 nm in diameter and has a linear positive-sense, single-stranded RNA genome [7]. The whole genome of this virus comprises a 5' untranslated region including a 5' leader sequence; an open reading frame (ORF) 1a/ab encoding nonstructural proteins (NSP) which are important for virus replication; four most important structural proteins of SARS-CoV-2 are the spike (S) protein- (trimeric) encoding gene, the envelope (E) protein-encoding gene, the membrane (M) protein-encoding gene and the nucleocapsid (N) protein-encoding gene; several accessory proteins such as ORF 3a, 6, 7a/b, and 8; and a 3' untranslated region. The ORF (1a/b), which is responsible for producing the two viral replicase polyproteins (PP1a and PP1ab). These replicase proteins are further processed into 16 different mature nonstructural proteins (NSPs), including NSP3 (papain-like protease), NSP5 (3C-like protease), NSP12 (RNA-dependent RNA polymerase [RdRp]), NSP13 (helicase), and other NSPs [1].

The S protein of SARS-CoV-2 is integrated into the viral envelope and considered the main key for the life cycle of viral entry into target host cells. For entry in human cells, it binds to cellular receptor angiotensin-converting enzyme 2 (ACE2). The S protein could be cleaved by host cellular protease furin into S1 and S2 subunits, which are responsible for recognition of host cell receptor and membrane fusion, respectively. The cleavage process of S protein plays a vital role in cell-cell mediated fusion which might exhibit increased cell-cell spread and potentially changed viral virulence. The presence of multibasic amino acids (multiple arginine) in a

S1/S2 cleavage site of SARS-CoV-2 is essential for supporting infection of human lung cells [8]. The S protein is first cleaved by a cellular endoprotease furin at the S1/S2 site in infected human cells, which is required for next the activity of the cellular serine protease TMPRSS2-mediated cleavage at the S2' site during viral entry into human lung cells. The S1/S2 domain in SARS-CoV-2 forms an exposed loop that carries multibasic amino acids that are found only in the human coronaviruses. Membrane fusion in SARS-CoV-2 and MERS-CoV depends on the cleavage of S protein by different host cell proteases at both the S1/S2 and the S2' sites [8].

The SARS-CoV-2 has 96% genomic identity with a previously detected SARS-like bat coronavirus [9]. Molecular characterization of nucleocapsid gene product from various coronavirus isolates, revealed that SARS-CoV 2 is more related to SARS-CoV 1 (more than 95% sequence similarity) and bat-CoV (92% sequence similarity). SARS-CoV 2 exhibited less resemblance with MERS-CoV (65% sequence similarity) [10]. In other study, the whole genome sequence of 2019-nCoV was closely related (89% nucleotide identity) to bat SARS-like-CoVZXC21 and (82% nucleotide identity) to human SARS-CoV. The phylogenetic analyses of orf1a/b, S, E, M and N genes also clustered with those of the civet and bat, and human SARS-CoV. However, the amino acid analysis showed that the identity of the external subdomain of spike's receptor-binding domain of 2019-nCoV shared only 40% amino acid identity with other SARS-related-CoV [11]. Pangolin-CoV isolated from a Malayan pangolin showed amino acid identity 100%, 98.6%, 97.8% and 90.7% with SARS-CoV-2 in the E, M, N and S proteins, respectively [12]. It was reported that the genome of the earlier SARS-Cov2/Wuhan isolates (2019-nCoV genome) had 88% nucleotide identity to two batSARS-like coronaviruses (bat-SL-CoVZC45 and bat-SL-CoVZXC21), but was more genetically distinct about 79% and 50% nucleotide identity from SARS-CoV and MERS-CoV, respectively. Phylogenetic analysis showed that 2019-nCoV belonged to the subgenus *Sarbecovirus* of β CoV, which is most closely related to bat-SL-CoVZC45 and bat-SL-CoVZXC21 [13].

This study aimed to conduct the molecular characterization and phylogenetic analyses of SARS-CoV-2 in Mediterranean Basin. The whole genome, Orf1ab gene, S gene, M gene, N gene and Orf3a gene sequences were included in the

study. In addition, the products of Orf1ab, S, M and N genes were also phylogenetically analyzed. Changes that occurred on the S-gene product of SARS-CoV-2 isolates recovered in Mediterranean basin countries were also detected.

2. MATERIALS AND METHODS

The whole genome sequence, Orf1ab gene sequence, S-gene sequence, M-gene sequence, N-gene sequence and Orf3a gene sequence of SARS-CoV-2 isolates recovered in Mediterranean basin countries were retrieved from GenBank Database previously available in National Center for Biotechnology Information (NCBI) using BLAST (Basic Local Alignment Search Tool) system. In addition, the product of Orf1ab gene, S-gene, M gene and N gene were also retrieved from GenBank Database for analyses. Multiple alignments were carried out using ClustalW of the computer program MEGA software (version 6) [14]. The Phylogenetic analyses for the cDNA of SARS-CoV-2 isolates recovered in Mediterranean basin countries were assessed by using Maximum Likelihood method based on the Tamura-Nei model [15]. Initial tree(s) for the heuristic searches were achieved by applying the Neighbor-Joining algorithm to a matrix of pairwise distances evaluated using the Maximum Composite Likelihood approach. However, for phylogenetic analyses of gene products, the pairwise distances were inferred by using the Maximum Likelihood method based on the Jones-Taylor Thornton (JTT) matrix-based model [16]. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances calculated using a JTT model. The robustness of the groupings in the Maximum Likelihood method was evaluated with 1000 bootstrap resampling. Reference sequences of other coronaviruses that retrieved from GenBank Database were also used for analyses. Accession numbers for all isolates that used to construct the phylogenetic analyses are illustrated on the trees.

Multiple alignments of 319 amino acid sequences of S gene product of SARS-Cov-2 isolates were conducted using ClustalW of the computer program MEGA software (version 6). Reference sequences for the S-gene product of other coronaviruses retrieved from GenBank Database. Multiple alignments of S-gene products were used to detect changes in S1/S2 cleavage site, S2' site and amino acid changes in whole product of that gene.

The identity between SARS-Cov-2 Wuhan-Hu-1 isolate (acc. no. MN908947.3) and other isolates such as batSARS-like-CoV isolates/bat-SL-CoVZC45 (acc. no. MG772933.1), bat-SL-CoVZXC21 (acc. no. MG772934.1), Pangolin-CoV isolates/PCoV_GX-P4L (acc. no. MT040333.1), PCoV_GX-P1E (acc. no. MT040334.1), PCoV_GX-P5L (acc. no. MT040335.1), PCoV_GX-P5E (acc. no. MT040336.1), PCoV_GX-P2V (acc. no. MT072864.1), and Pangolin-CoV MP789 (acc. no. MT121216.1) and Bat-CoV RaTG13 isolate (Acc. no. MN996532.1) was also included.

Accession numbers for SARS-CoV-2 isolates (Accessed July 20, 2020) recovered in Mediterranean basin countries that retrieved from GenBank Database and used in this study are Israel (2 isolates: MT276598.1, MT276597.1); Morocco (10 isolates: MT513758.1, MT568645.1, MT731285.1, MT731292.1, MT731327.1, MT731468.1, MT731673.1, MT731746.1, MT731764.1, MT733120.1); Tunisia (8 isolates: MT499215.1-MT499220.1, MT559038.1, MT324682.1); Turkey (7 isolates: MT560525.1, MT560530.1, MT560531.1, MT605818.1, MT675956.1, MT675958.1, MT327745.1); Spain (27 isolates: MT359865.1, MT359866.1, MT576025.1-MT576030.1, MT198652.2, MT233519.1, MT233521.1, MT233523.1, MT292569.1-MT292577, MT292579.1, MT655131.1- MT655135.1); Egypt (74 isolates: MT510690.1- MT510703.1, MT511059.1, MT511065.1, MT511083.1, MT511084.1, MT517443.1, MT658507.1- MT658509.1, MT658264.1, MT658500.1-MT658505.1, MT661524.1, MT609918.1, MT610913.1, MT611527.1, MT611530.1, MT611536.1-MT611539.1, MT611440.1-MT611444.1, MT611448.1-MT611451.1, MT611968.1, MT614347.1-MT614349.1, MT614356.1, MT614357.1, MT614595.1-MT614601.1, MT624728.1, MT624733.1, MT627389.1-MT627392.1, MT627394.1-MT627396.1, MT627404.1-MT627406.1, MT627393.1, MT648830.1); France (81 isolates: MT470100.1-MT470179.1, MT320538.2); Greece (97 isolates: MT328032.1-MT328035.1, MT459832.1-MT459851.1, MT459853.1- MT459925.1) and Italy (13 isolates: MT126808.1, MT531537.2, MT622321.1, MT682732.1, MT525950.1, MT527178.1, MT527184.1, MT528235.1, MT528237.1-MT528239.1, MT066156.1, MT077125.1).

3. RESULTS

Results of the current study reported that the whole genome sequence of the earlier SARS-Cov-2 Wuhan-Hu-1 isolate (acc. no. MN908947.3) had 89.12% and 88.65% nucleotide identity for batSARS-like-Cov/bat-SL-CoVZC45 isolate (acc. no. MG772933.1) and bat-SL-CoVZXC21 isolate (acc. no. MG772934.1), respectively. Pangolin-CoV isolates PCoV_GX-P4L (acc. no. MT040333.1), PCoV_GX-P1E (acc. no. MT040334.1), PCoV_GX-P5L (acc. no. MT040335.1), PCoV_GX-P5E (acc. no. MT040336.1), PCoV_GX-P2V (acc. no. MT072864.1), and Pangolin-CoV MP789 (acc. no. MT121216.1) had nucleotide identity 85.97%, 85.95, 85.98%, 85.94%, 85.94% and 90.11%, respectively, with the whole genome of SARS-Cov2/Wuhan-Hu-1 isolate. Analysis of the whole genome of the SARS-Cov2/Wuhan-Hu-1 isolate showed that it was most closely related to Bat-CoV RaTG13 isolate (Acc. no. MN996532.1) than other coronaviruses deposited in GenBank. The nucleotide identity between the genome of the SARS-Cov2/Wuhan-Hu-1 isolate and Bat-CoV RaTG13 isolate is 96.12%. Bat-CoV RaTG13 isolate was recovered from a fecal swab from a host *Rhinolophus affinis* in China, on July, 2013. In addition, the whole genome of SARS-Cov2/Wuhan-Hu-1 isolate is closely related to

the Pangolin-CoV isolate MP789 (nucleotide identity 90.11%). This isolate was recovered in China on March, 2019 from the lung of *Manis javanica*. The amino acid sequence identity was carried out between genes (Orf1ab, S, M and N) of batSARS-like-CoV/bat-SL-CoVZC45, batSARS-like-CoV/bat-SL-CoVZXC21, Pangolin-CoV isolates and Bat-CoV RaTG13 isolate and earlier SARS-Cov2/Wuhan-Hu-1 isolate. Results of amino acid identity are shown in Table 1.

In the current study the whole genome sequence, Orf1ab gene sequence, M-gene sequence, N-gene sequence, Orf3a gene sequence and S-gene sequence of SARS-CoV-2 isolates recovered in Mediterranean basin countries (France, Spain, Italy, Egypt, Greece, Italy, Morocco, Tunisia, Turkey and Israel) retrieved from GenBank Database are used for phylogenetic analyses. Other reference sequences of different coronavirus strains retrieved from GenBank Database are also included in this study. Phylogenetic analyses showed that these isolates clustered with SARS-Cov2/Wuhan-Hu-1 isolate. In addition, phylogenetic analyses showed that SARS-Cov2/Wuhan-Hu-1 isolate and isolates recovered in Mediterranean basin countries are clustered to Bat-CoV RaTG13 isolate using the whole genome, Orf1ab gene sequence, S gene sequence, M gene sequence, N gene sequence

and Orf3a gene sequence. Results are presented in Fig. 1A, 1B, 1C, 1D, 1E and 1F. According to the Orf1ab, S, M and N gene products, SARS-CoV-2 isolates are clustered with Bat-CoV RaTG13 isolate Fig. 2A, 2B, 2C and 2D. Results of this study showed that Bat-CoV RaTG13 isolate is more closely related to SARS-CoV-2 isolates than Pangolin-CoV isolate MP789 or other Pangolin-CoV isolates Table 1, Fig. 1 and 2.

Analyses of 319 S-gene product of SARS-Cov-2 isolates recovered in Mediterranean basin countries, showed that all these isolates harbored a highly cleavable S1/S2 cleavage site, which is not found in other closely related coronaviruses Figs. 3 and 4. Basic amino acids in The S1/S2 cleavage sites of some these isolates are shaded in dark gray Fig 3. In addition, Histidine (H) is present instead of Glutamine (Q) in 8/74 (10.8%) isolates recovered in Egypt, while Isoleucine (I) is present instead of Threonine (T) in 2/81 (2.47%) isolates recovered in France. Histidine and Isoleucine in these isolates are shaded in light gray Fig. 3. In addition, the DPSKRSEFIEDLLFNKV protein sequence motif was seen in all S gene products of SARS-CoV-2 isolates recovered in Mediterranean basin countries Fig. 3.

The most and less closely related coronavirus strains Bat-CoV RaTG13 isolate and Pangolin-CoV isolate MP789, respectively, have similar S1/S2 motif sequence to SARS-CoV-2. The SARS-CoV-2 has the multibasic cleavage sequence "SYQTQTNSPRRARSVA", while Bat-CoV RaTG13 and Pangolin-CoV MP789 strains have the monbasic cleavage sequence "SYQTQTNS----RSVS". Sequences of S1/S2 motif for Bat-CoV RaTG13 and Pangolin-CoV MP789 strains are presented in Fig. 4 (shaded in light gray).

Analyses of S gene and its product, showed that only one isolate out 319 had a deletion mutation. The Egyptian isolate which has an accession number (MT658264.1) had 430TATdel in surface glycoprotein S gene compared to the reference isolate SARS-Cov-2 Wuhan-Hu-1 isolate (acc. no. MN908947.3). This mutation leads to 145Ydel in S protein (QKX64992.1) Table 2. The most dominant mutation at the level of S gene product was 614 D→G. The frequency of this

mutation among SARS-CoV-2 isolates recovered in Mediterranean basin countries had a range 48.1% -100% with a total frequency 253 out 329 (79.3%) isolates Table 2.

4. DISCUSSION

Analysis of the whole genome of the SARS-CoV-2 Wuhan-Hu-1 isolate showed that it was closely related (96.12% nucleotide identity) to Bat-CoV RaTG13 isolate than other coronaviruses deposited in GenBank. On the other hand, Pangolin-CoV MP789 isolate was less related (nucleotide identity 90.11%) than Bat-CoV RaTG13 isolate. According to the phylogenetic analyses this study showed, that SARS-CoV-2 isolates recovered in Mediterranean basin countries were similar to the reference SARS-Cov-2 Wuhan-Hu-1 isolate and all of them were clustered very closely with Bat-CoV RaTG13 isolate and less closely with Pangolin-CoV MP789 isolate. These results are in agreement with report published previously, which showed that the likely reservoir hosts for SARS-CoV-2 are bats [1]. In the current study, phylogenetic analyses showed that Bat-CoV RaTG13 considered as the ancestor of SARS-CoV-2. *Rhinolophus affinis* is considered a possible host in the emergence of SARS-CoV-2, that could infect humans directly or indirectly by transmitting the virus into intermediate hosts that facilitate the transmission process from animal-to-humans. Previous studies based on metagenomic sequencing for the samples from Malayan pangolins (*Manis javanica*) in China, identified pangolin-associated coronaviruses that belong to two sub-lineages of SARS-CoV-2-related coronaviruses [12,17]. Xiao *et al.* [12] reported that pangolin coronavirus genomes have 85.5% to 92.4% sequence identity to SARS-CoV-2 and SARS-CoV-2 is closely related to one of these sub-lineages (GD/P1L and GD/P2S). They also considered the Malayan pangolins as possible hosts in the emergence of new coronaviruses, but they didn't exclude that pangolins acquired their SARS-CoV-2-related coronaviruses independently from bats or other animals. This suggests that Pangolins may be one of the hosts for these viruses and the SARS-CoV-2 may have originated in the recombination of a virus similar to pangolin-CoV with one similar to Bat-CoV RaTG13 [12].

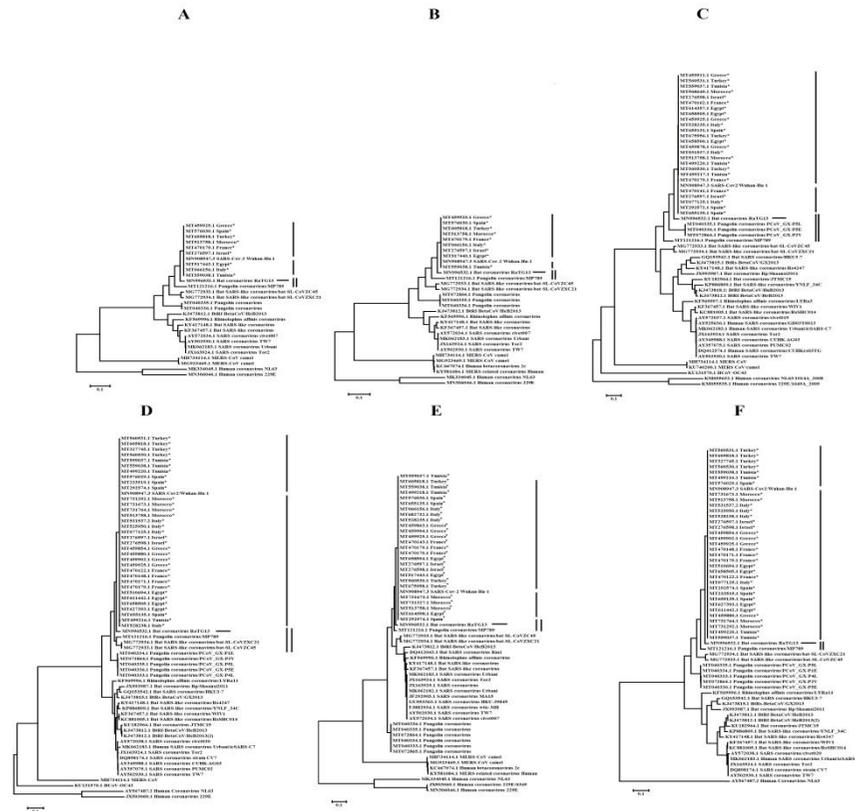


Fig. 1. Phylogenetic analyses by Maximum Likelihood method based on the whole genome sequence (A), Orf1ab gene sequence (B), S gene sequence (C), M gene sequence (D), N gene sequence (E) and Orf3a gene sequence (F) of human SARS-CoV-2 recovered in Mediterranean basin countries retrieved from GenBank Database (denoted by asterisk). Reference sequences of other coronaviruses retrieved from GenBank Database were used for phylogenetic analyses. Accession numbers for all isolates are illustrated on the tree. Phylogenetic analyses were based on alignments obtained from ClustalW of 25649 positions, 19492 positions, 3132 positions, 644 positions, 1056 positions and 669 positions for the whole genome sequence, Orf1ab gene sequence, S gene sequence, M gene sequence, N gene sequence and Orf3a gene sequence, respectively. The robustness of the groupings in the Maximum Likelihood method analysis was evaluated with 1000 bootstrap resamplings. The SARS-CoV-2 Wuhan-Hu-1 isolate (acc. no. MN908947.3) is considered the major reference for SARS-CoV-2

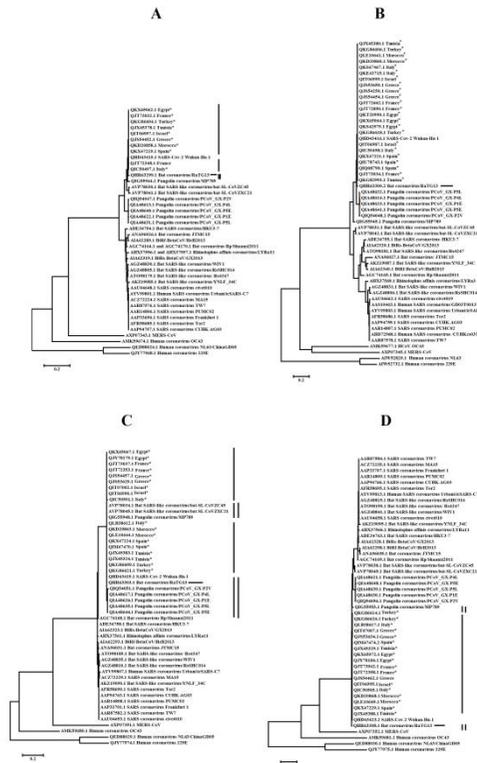


Fig. 2. Phylogenetic analyses by Maximum Likelihood method based on the sequence of Orf1ab gene product (A), S gene product (B), M gene product (C) and N gene product (D) of human SARS-CoV-2 recovered in Mediterranean basin countries retrieved from GenBank Database (denoted by asterisk). Reference sequences of other coronaviruses retrieved from GenBank Database were used for phylogenetic analyses. Accession numbers for all isolates are illustrated on the tree. Phylogenetic analyses were based on alignments obtained from ClustalW of 6137 positions, 1035 positions, 149 positions, 348 positions for the Orf1ab gene product sequence, S gene product sequence, M gene product sequence and N gene product sequence, respectively. The robustness of the groupings in the Maximum Likelihood method analysis was evaluated with 1000 bootstrap resamplings. The gene product of SARS-CoV-2 Wuhan-Hu-1 isolate (acc. no. MN908947.3) is considered the major reference for SARS-CoV-2

Accession number/Country	S1/S2	Accession number/Country	S2'
QKG86606.1 Turkey	S Y Q T Q T N S P A S V A	QKG86606.1 Turkey	D P S P S K S F I E D L L F N K V
QLA09894.1 Turkey	S Y Q T Q T N S P A S V A	QLA09894.1 Turkey	D P S P S K S F I E D L L F N K V
QJX45380.1 Tunisia	S Y Q T Q T N S P A S V A	QJX45380.1 Tunisia	D P S P S K S F I E D L L F N K V
QJX45356.1 Tunisia	S Y Q T Q T N S P A S V A	QJX45356.1 Tunisia	D P S P S K S F I E D L L F N K V
QJX45333.1 Tunisia	S Y Q T Q T N S P A S V A	QJX45333.1 Tunisia	D P S P S K S F I E D L L F N K V
QKX47221.1 Spain	S Y Q T Q T N S P A S V A	QKX47221.1 Spain	D P S P S K S F I E D L L F N K V
QIQ08810.1 Spain	S Y Q T Q T N S P A S V A	QIQ08810.1 Spain	D P S P S K S F I E D L L F N K V
QKJ68400.1 Spain	S Y Q T Q T N S P A S V A	QKJ68400.1 Spain	D P S P S K S F I E D L L F N K V
QKD20860.1 Morocco	S Y Q T Q T N S P A S V A	QKD20860.1 Morocco	D P S P S K S F I E D L L F N K V
QKI36913.1 Morocco	S Y Q T Q T N S P A S V A	QKI36913.1 Morocco	D P S P S K S F I E D L L F N K V
QIT06999.1 Israel	S Y Q T Q T N S P A S V A	QIT06999.1 Israel	D P S P S K S F I E D L L F N K V
QIT06987.1 Israel	S Y Q T Q T N S P A S V A	QIT06987.1 Israel	D P S P S K S F I E D L L F N K V
QKX65040.1 Egypt	S Y Q T H T N S P A S V A	QKX65040.1 Egypt	D P S P S K S F I E D L L F N K V
QKS66928.1 Egypt	S Y Q T H T N S P A S V A	QKS66928.1 Egypt	D P S P S K S F I E D L L F N K V
QKS66904.1 Egypt	S Y Q T H T N S P A S V A	QKS66904.1 Egypt	D P S P S K S F I E D L L F N K V
QKS66892.1 Egypt	S Y Q T H T N S P A S V A	QKS66892.1 Egypt	D P S P S K S F I E D L L F N K V
QKS42609.1 Egypt	S Y Q T H T N S P A S V A	QKS42609.1 Egypt	D P S P S K S F I E D L L F N K V
QKR84285.1 Egypt	S Y Q T H T N S P A S V A	QKR84285.1 Egypt	D P S P S K S F I E D L L F N K V
QKR84225.1 Egypt	S Y Q T H T N S P A S V A	QKR84225.1 Egypt	D P S P S K S F I E D L L F N K V
QKX65100.1 Egypt	S Y Q T H T N S P A S V A	QKX65100.1 Egypt	D P S P S K S F I E D L L F N K V
QKX65052.1 Egypt	S Y Q T Q T N S P A S V A	QKX65052.1 Egypt	D P S P S K S F I E D L L F N K V
QKX64992.1 Egypt	S Y Q T Q T N S P A S V A	QKX64992.1 Egypt	D P S P S K S F I E D L L F N K V
QKT21062.1 Egypt	S Y Q T Q T N S P A S V A	QKT21062.1 Egypt	D P S P S K S F I E D L L F N K V
QJT72902.1 France	S Y Q I Q T N S P A S V A	QJT72902.1 France	D P S P S K S F I E D L L F N K V
QJT72614.1 France	S Y Q I Q T N S P A S V A	QJT72614.1 France	D P S P S K S F I E D L L F N K V
QJT72974.1 France	S Y Q T Q T N S P A S V A	QJT72974.1 France	D P S P S K S F I E D L L F N K V
QJT73034.1 France	S Y Q T Q T N S P A S V A	QJT73034.1 France	D P S P S K S F I E D L L F N K V
QJT72566.1 France	S Y Q T Q T N S P A S V A	QJT72566.1 France	D P S P S K S F I E D L L F N K V
QJS54118.1 Greece	S Y Q T Q T N S P A S V A	QJS54118.1 Greece	D P S P S K S F I E D L L F N K V
QJS53986.1 Greece	S Y Q T Q T N S P A S V A	QJS53986.1 Greece	D P S P S K S F I E D L L F N K V
QJS53710.1 Greece	S Y Q T Q T N S P A S V A	QJS53710.1 Greece	D P S P S K S F I E D L L F N K V
QIC50498.1 Italy	S Y Q T Q T N S P A S V A	QIC50498.1 Italy	D P S P S K S F I E D L L F N K V
QKE43679.1 Italy	S Y Q T Q T N S P A S V A	QKE43679.1 Italy	D P S P S K S F I E D L L F N K V

Fig. 3. The alignment of amino acid sequences around the highly cleavable S1/S2 and S2' cleavage sites of SARS-Cov-2 strains recovered in Mediterranean basin countries (France, Spain, Italy, Egypt, Greece, Italy, Morocco, Tunisia, Turkey and Israel) that retrieved from GenBank Database. Basic amino acids in the cleavage sites of some these isolates are shaded in dark gray, while mutant amino acids in the S1/S2 cleavage site are shaded in light gray

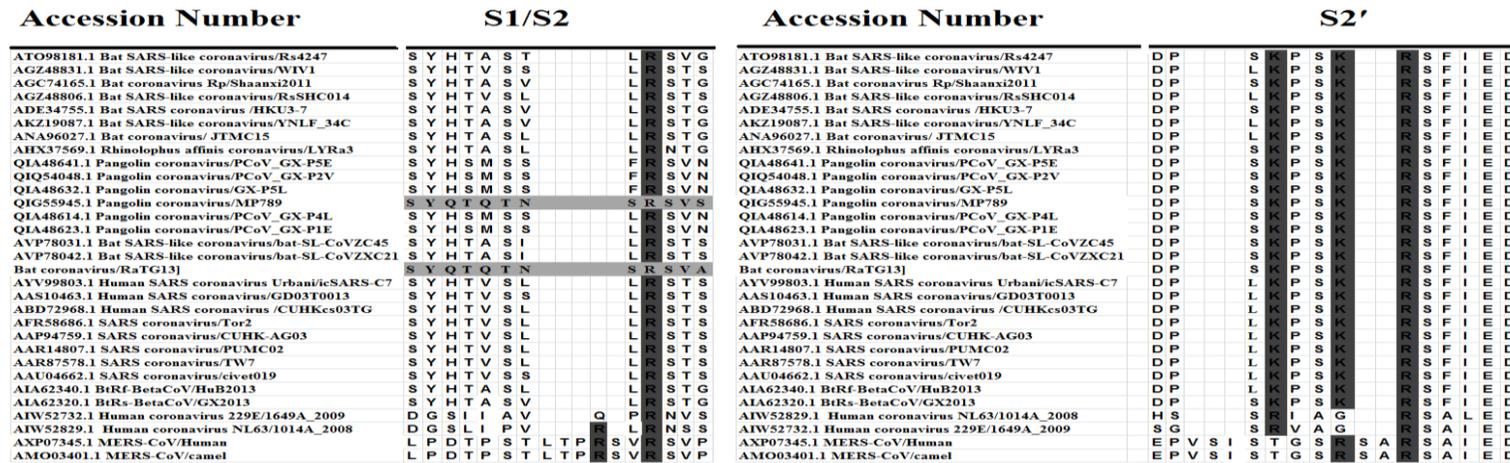


Fig. 4. The alignment of amino acid sequences around the S1/S2 and S2' cleavage sites of coronavirus strains found in bats, pangolin, humans, civet cats and that can cause infection in humans. Amino acid sequences of these coronavirus strains are retrieved from GenBank Database. Basic amino acids in the cleavage sites of these isolates are shaded in dark gray. The S1/S2 motif sequence of Bat-CoV RaTG13 and Pangolin-CoV MP789 isolates are shaded in light gray

Table 1. The amino acid identity between the gene product (Orf1ab, S, M and N gene) of the SARS-CoV-2 Wuhan-Hu-1 and other closely related coronavirus isolates

Coronavirus Strain (acc. no.)	Amino acid identity %			
	Orf1ab	S	M	N
batSARS-like-CoV/bat-SL-CoVZC45 (MG772933.1)	95.55%	81%	98.65%	94.27%
batSARS-like-CoV/bat-SL-CoVZXC21 (MG772934.1)	95.15%	80.32%	98.65%	94.27%
Pangolin-CoV/PCoV_GX-P4L (MT040333.1)	92.59%	92.3%	98.2%	93.79%
Pangolin-CoV/PCoV_GX-P1E (MT040334.1)	92.58%	92.22%	98.2%	94.03%
Pangolin-CoV/PCoV_GX-P5L (MT040335.1)	92.58%	92.38%	98.2%	93.79%
Pangolin-CoV/PCoV_GX-P5E (MT040336.1)	92.55%	92.3%	98.2%	93.32%
Pangolin-CoV/PCoV_GX-P2V (MT072864.1)	92.49%	92.14%	97.75%	93.79%
Pangolin-CoV/MP789 (MT121216.1)	96.68%	90.5%	98.2%	97.85%
Bat-CoV RaTG13 (MN996532.1)	98.55%	97.41%	99.55%	99.05%

Table 2. Changes in S gene product of SARS-Cov-2 isolates recovered in Mediterranean basin countries compared to SARS-CoV-2 reference isolate Wuhan-Hu-1 (acc. no. QHD43416.1) retrieved from GenBank

Country	Accession number of S gene product	Site of mutation
Turkey	QLA09870.1	188 N→K
	QKG86606.1, QKG86594.1, QLA09894.1, QLA09870.1, QKQ11725.1, QKG86618.1	614 D→G
	QIZ16509.1	772 V→I
Tunisia	QJX45356.1	29 T→I
	QJX45333.1	279 Y→N
	QJX45344.1	288 A→T
	QJX45380.1, QJX45368.1, QJX45356.1, QJX45344.1, QKG82981.1	614 D→G
Morocco	QKD20860.1, QKI36913.1, QLF67497.1, QLE11089.1, QLE11077.1, QLE10921.1, QLE10668.1, QLE10641.1, QLE10629.1, QLE00003.1	614 D→G
Israel	QIT06987.1	614 D→G
Italy	QKE43691.1	5 L→F
	QKE43727.1, QKE43715.1, QKE43703.1, QKE43691.1, QKE43679.1	614 D→G
	QKE43667.1, QKE32976.1, QLB38609.1, QKS67467.1, QKE45430.1	
	QLB38609.1	1260 D→H
Spain	QKJ68364.1	213 V→L
	QKX47197.1, QKX47173.1, QIU78779.1, QIU78719.1, QIU78707.1, QKJ68400.1, QKJ68388.1, QKJ68376.1, QKJ68364.1, QKJ68352.1, QKJ68340.1, QJC21017.1, QJC21005.1	614 D→G
Egypt	QKX65028.1, QKX64992.1	5 L→F
	QKR84285.1	12 S→F
	QKR84321.1	57 A→S
	QKX64992.1	145 Ydel
	QKX65076.1	242 L→F
	QKT21014.1	408 R→Y
	QJY78056.1	408 R→I
	QKX65064.1, QKX65052.1, QKX65040.1, QKX65028.1, QKX65016.1, QKX65004.1, QKX64992.1, QKW95051.1, QKT20990.1, QKT21050.1, QKT21062.1, QKT21038.1, QKT21026.1, QKT21014.1, QKT21002.1, QKT20978.1, QKT20966.1, QKT20954.1, QKT20942.1, QKS74806.1, QKS74794.1, QKS66928.1, QKS66916.1, QKS66904.1, QKS66892.1, QKS66880.1, QKS66868.1, QKS42633.1, QKS42621.1, QKS42609.1, QKS42591.1, QKS42579.1, QKR84393.1, QKR84381.1, QKR84369.1, QKR84357.1, QKR84345.1, QKR84333.1, QKR84321.1, QKR84285.1, QKR84297.1, QKR84309.1, QKR84273.1, QKR84261.1, QKR84249.1, QKR84237.1, QKR84225.1, QKR83985.1, QKQ63927.1,	614 D→G

Country	Accession number of S gene product	Site of mutation
France	QKX96275.1, QKX65100.1, QKX65088.1, QKX65076.1, QJY78068.1, QJZ28359.1, QJZ28347.1, QJZ28126.1, QJZ28114.1, QJY78176.1, QJY78164.1, QJY78152.1, QJY78140.1, QJY78104.1, QJY78128.1, QJY78116.1, QJY78092.1, QJY78080.1, QJY78056.1, QJY78044.1, QJY78020.1, QJY78032.1, QKW95051.1	640 S→A
	QJZ28114.1	653 A→V
	QKX65052.1, QKS66928.1, QKS66904.1, QKS66892.1, QKR84285.1, QKR84225.1, QKX65100.1	677 Q→H
	QJT72722.1, QJT72410.1, QJT72134.1	5 L→F
	QJT72086.1	153 M→I
	QJT72350.1	167 L→I
	QJT73010.1	240 T→I
	QJT72806.1	382 V→E
	QJT72386.1	379 C→F
	QJT73034.1	553 T→N
	QJT72470.1	572 T→I
	QJT72278.1	611 L→F
	QJT73022.1, QJT73010.1, QJT72998.1, QJT72962.1, QJT72950.1, QJT72938.1, QJT72926.1, QJT72914.1, QJT72902.1, QJT72890.1, QJT72878.1, QJT72866.1, QJT72854.1, QJT72830.1, QJT72818.1, QJT72806.1, QJT72770.1, QJT72758.1, QJT72746.1, QJT72734.1, QJT72722.1, QJT72698.1, QJT72674.1, QJT72662.1, QJT72650.1, QJT72638.1, QJT72614.1, QJT72602.1, QJT72566.1, QJT72542.1, QJT72530.1, QJT72518.1, QJT72506.1, QJT72494.1, QJT72482.1,	614 D→G
	QJT72470.1, QJT72422.1, QJT72410.1, QJT72398.1, QJT72386.1,	
	QJT72374.1, QJT72362.1, QJT72338.1, QJT72326.1, QJT72314.1, QJT72302.1, QJT72290.1, QJT72278.1, QJT72266.1, QJT72254.1,	
	QJT72242.1, QJT72230.1, QJT72206.1, QJT72194.1, QJT72170.1,	
	QJT72158.1, QJT72146.1, QJT72122.1, QJT72110.1, QJT72098.1,	
	QIX12148.2	
	QJT72902.1, QJT72614.1	676 T→I
	QJT72794.1, QJT72710.1, QJT72626.1, QJT72554.1, QJT72350.1, QJT72182.1, QJT72086.1	845 A→S
QJT72242.1	1020 A→V	
QJT72902.1, QJT72614.1	1162 P→L	
QJT72386.1	1263 P→L	
Greece	QJS53410.1	188 N→D
	QJS53494.1, QIZ16559.1	197 I→V
	QJS54106.1	393 T→P

Country	Accession number of S gene product	Site of mutation
	QJS54454.1, QJS54418.1, QJS54406.1, QJS54394.1, QJS54358.1, QJS54346.1, QJS54334.1, QJS54322.1, QJS54310.1, QJS54298.1, QJS54286.1, QJS54274.1, QJS54262.1, QJS54250.1, QJS54238.1, QJS54226.1, QJS54214.1, QJS54202.1, QJS54190.1, QJS54178.1, QJS54166.1, QJS54154.1, QJS54130.1, QJS54118.1, QJS54106.1, QJS54094.1, QJS54070.1, QJS54058.1, QJS54046.1, QJS54034.1, QJS54010.1, QJS53998.1, QJS53986.1, QJS53974.1, QJS53962.1, QJS53914.1, QJS53902.1, QJS53890.1, QJS53878.1, QJS53866.1, QJS53854.1, QJS53842.1, QJS53830.1, QJS53818.1, QJS53806.1, QJS53794.1, QJS53782.1, QJS53770.1, QJS53746.1, QJS53710.1, QJS53662.1, QJS53650.1, QJS53638.1, QJS53626.1, QJS53614.1, QJS53602.1, QJS53590.1, QJS53566.1, QJS53542.1, QJS53530.1, QJS53506.1, QJS53470.1, QJS53458.1, QJS53446.1, QJS53434.1, QJS53422.1, QJS53410.1, QJS53386.1, QIZ16547.1, QIZ16535.1, QJS53374.1, QJS53362.1, QJS53350.1, QJS53338.1, QIZ16571.1, QIZ16559.1	614 D→G
	QJS54286.1	769 G→V
	QJS53386.1	789 Y→D
	QJS53506.1, QIZ16559.1	1101 H→Y
	QJS53398.1	1122 V→L
	QJS53506.1	1191 K→N
	QJS54346.1, QJS54334.1, QJS53818.1	1263 P→L

The existence of several basic amino acid residues at the S1/S2 and S2' cleavage sites is necessary for an effective SARS-CoV-2 S protein proteolytic cleavage in host human lung cells. The cleavage process at these sites by different cellular host protease is essential for S protein-mediated cell-cell fusion and entry into human lung cells. Thus, drugs or chemicals that used to block or inhibit the action of proteolytic processing of these enzymes might be considered as an optional treatment for SARS-CoV-2 infection [8]. Viruses that have a monobasic cleavage site are digested by certain proteases such as TMPRSS2 with an infection profile restricted to the respiratory tract and results in moderate or mild disease. While, the viruses with a multibasic cleavage site are activated by everywhere expressed furin and related proprotein convertases, and can thus spread these viruses systemically and cause severe infection and high mortality rates in infected individuals. In the context of SARS-CoV-2 infection, the existence of a highly cleavable S1/S2 cleavage site in S-gene product may therefore not have been unexpected. However, it is noticeable that all SARS-CoV-2-related coronaviruses of bats and pangolins recognized today are activated at a monobasic cleavage site [8]. The most and less closely related coronavirus strains Bat-CoV RaTG13 isolate and Pangolin-CoV isolate MP789, respectively, have similar but not identical S1/S2 motif sequence to SARS-CoV-2. This may indicate that Bat-CoV RaTG13 which is considered the possible origin of SARS-CoV-2 acquired the multibasic S1/S2 motif by recombination or insertion. During the alignment of Orf1ab gene product, we observed that this gene encoded the sequence PRR. The PRR sequence was determined at positions 5910 and 5916 for Pangolin-CoV MP789 isolate and Bat-CoV RaTG13 isolate, respectively. Many other coronavirus strains that used in this study showed that they have PRR sequence at Orf1ab gene product. This may support multibasic S1/S2 motif acquired by recombination between different genes. Theoretically, these strains may become human pathogenic at any time they acquired this sequence.

The DPSKRSFIEDLLFNKV protein sequence motif was seen in all S gene products of SARS-CoV-2 isolates recovered in Mediterranean basin countries. This motif is required for proteolytic activation cleavage, and has a greater degree of residue conservation on the surface of the virus, suggesting that mutations in this sequence motif are very rare and much less easily. This protein

sequence motif appears to be a potential primary target for various ongoing synthetic vaccine efforts and a basis for drug development [18].

The most dominant mutation at the level of S gene product was 614 D→G. The frequency of this mutation among SARS-CoV-2 isolates recovered in Mediterranean basin countries was 253 out 329 (79.3%) isolates. This mutation is located in S1 and may play a role in transmission and pathogenicity of this virus. The impact of 145Ydel, 614 D→G and other mutations detected in S-gene product on virus transmission, pathogenicity, diagnosis and strategies of antiviral therapy, should be rapidly assessed in further studies. This study suggests that these mutations play a role in virulence and infection of SARS-CoV-2.

5. CONCLUSION

Phylogenetic analyses of the sequences analyzed in this study, showed that SARS-CoV-2 isolates are more closely related to Bat-CoV RaTG13 Isolate than Pangolin-CoV isolates. The S gene product of this virus mediates entry into the human host cell and has S1/S2 cleavage site containing multibasic amino acids (PRRAR) which are not found in other closely related coronaviruses. Many coronavirus strains that were deposited in GenBank, showed that they have PRR sequence in Orf1ab gene product. This may strongly support that multibasic S1/S2 motif acquired by a process of recombination or insertion. Theoretically, any coronavirus strain acquired this sequence becomes highly pathogenic to humans. The dominant mutation (79.3%) at S-gene product level was 614D→G. The impact of mutations detected in S gene product on virus transmission, diagnosis, pathogenicity and strategies of antiviral therapy, should be rapidly assessed in further studies.

ACKNOWLEDGEMENT

The author thanks Prof. Saleh A. Naser, University of Central Florida, College of Medicine, for language revision.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Liu YC, Kuo RL, Shih SR. COVID-19: The first documented coronavirus pandemic in history. *Biomed J.* 2020;43(4):328-333.
2. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, et al. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Med Infect Dis.* 2020;34(101623):1-13.
3. Chan JF, Yuan S, Kok KH, To KKW, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet.* 2020;395(10223):514-523.
4. Bastola A, Sah R, Rodriguez-Morales AJ, Lal BK, Jha R, Ojha HC, et al. The first 2019 novel coronavirus case in Nepal. *Lancet Infect Dis.* 2020;20(3):279-280.
5. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* 2020;579(7798):265-269.
6. Prajapat M, Sarma P, Shekhar N, Avti P, Sinha S, Kaur H, et al. Drug targets for corona virus: A systematic review. *Indian J Pharmacol.* 2020;52(1):56-65.
7. Bar-On YM, Flamholz A, Phillips R, Milo R. SARS-CoV-2 (COVID-19) by the numbers. *Elife.* 2020;9:(e57309).
8. Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell.* 2020;78(4):779-784.
9. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020;579(798):270-273.
10. Kannan S, Subbaram K, Ali S, Kannan H. Molecular characterization and amino acid homology of nucleocapsid (N) protein in SARS-CoV-1, SARS-CoV-2, MERS-CoV, and bat coronavirus. *J Pure Appl Microbiol.* 2020;14(suppl. 1):757-763.
11. Chan JFW, Kok KH, Zhu Z, Chu H, To KKW, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect.* 2020;9(1):221-236.
12. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou JJ, et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature.* 2020;583(7815):286-289.
13. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet.* 2020;395(10224):565-574.
14. Tamura K, Stecher G, Peterson D, FilipSKI A and Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725-2729.
15. Tamura K and Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 1993.10(3):512-526.
16. Jones, DT, Taylor WR and Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 1992;8(3):275-282.
17. Lam TT, Jia N, Zhang YW, Shum MH, Jiang JF, Zhu HC, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature.* 2020;583(7815):282-285.
18. Robson B. COVID-19 coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance. *Comput Biol Med.* 2020;121:103749.

© 2021 Adwan; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:
<http://www.sdiarticle4.com/review-history/67086>