Research article

Temporal increase in D614G mutation of SARS-CoV-2 in the Middle East and North Africa

Malik Sallam [a,b,c,*], Nidaa A. Ababneh [d], Deema Dababseh [e], Faris G. Bakri [f,g,h], Azmi Mahafzah [a,b]

[a] Department of Pathology, Microbiology and Forensic Medicine, School of Medicine, The University of Jordan, Amman, Jordan
[b] Department of Clinical Laboratories and Forensic Medicine, Jordan University Hospital, Amman, Jordan
[c] Department of Translational Medicine, Faculty of Medicine, Lund University, Malmö, Sweden
[d] Cell Therapy Center (CTC), The University of Jordan, Amman, Jordan
[e] School of Dentistry, The University of Jordan, Amman, Jordan
[f] Department of Internal Medicine, School of Medicine, The University of Jordan, Amman, Jordan
[g] Department of Internal Medicine, Jordan University Hospital, Amman, Jordan
[h] Infectious Diseases and Vaccine Center, University of Jordan, Amman, Jordan

ARTICLE INFO

ABSTRACT

Background: Phylogeny construction can help to reveal evolutionary relatedness among molecular sequences. The spike (*S*) gene of SARS-CoV-2 is the subject of an immune selective pressure which increases the variability in such region. This study aimed to identify mutations in the *S* gene among SARS-CoV-2 sequences collected in the Middle East and North Africa (MENA), focusing on the D614G mutation, that has a presumed fitness advantage. Another aim was to analyze the *S* gene sequences phylogenetically.
Methods: The SARS-CoV-2 *S* gene sequences collected in the MENA were retrieved from the GISAID public database, together with its metadata. Mutation analysis was conducted in Molecular Evolutionary Genetics Analysis software. Phylogenetic analysis was done using maximum likelihood (ML) and Bayesian methods.
Result: A total of 553 MENA sequences were analyzed and the most frequent *S* gene mutations included: D614G = 435, Q677H = 8, and V6F = 5. A significant increase in the proportion of D614G was noticed from (63.0%) in February 2020, to (98.5%) in June 2020 ($p < 0.001$). Two large phylogenetic clusters were identified via ML analysis, which showed an evidence of inter-country mixing of sequences, which dated back to February 8, 2020 and March 15, 2020 (median estimates). The mean evolutionary rate for SARS-CoV-2 was about 6.5×10^{-3} substitutions/site/year based on large clusters' Bayesian analyses.
Conclusions: The D614G mutation appeared to be taking over the COVID-19 infections in the MENA. Bayesian analysis suggested that SARS-CoV-2 might have been circulating in MENA earlier than previously reported.

1. Introduction

Members of *Coronaviridae* family of viruses have started to gain a substantial interest due to their potential role as causative agents of emerging infections in humans (Fehr and Perlman, 2015). This was manifested by the 2002–2003 SARS outbreak, 2012 MERS outbreak, and the current coronavirus disease 2019 (COVID-19) pandemic, the first documented coronavirus pandemic, which can be viewed as the full-blown consequence of coronavirus threat (Cherry and Krogstad, 2004; Liu et al., 2020; Lu and Liu, 2012; Peiris et al., 2003).

The causative agent of COVID-19 is severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with a relatively high mutation rate that is mostly related to its RNA-dependent RNA polymerase, with minimal proofreading activity (Duffy et al., 2008; Liu et al., 2020; Sevajol et al., 2014). In addition, the high frequency of recombination in coronaviruses augments its genetic diversity and its ability of cross-species transmission (Su et al., 2016; Woo et al., 2009). The aforementioned features are accompanied by ubiquitous presence of coronaviruses in various animal reservoirs (Guan et al., 2003). Thus, cross-species transmission, including spread to humans seems an inevitable outcome (Graham and Baric, 2010; Woo et al., 2009). This is mainly related to

human, ecologic and economic factors, which explain the increased frequency of zoonosis (Delabouglise et al., 2017; Karesh et al., 2012; Morse et al., 2012).

The spike glycoprotein of SARS-CoV-2, is responsible for attachment of the virus to its cellular receptor (angiotensin-converting enzyme 2 [ACE2]) (Fehr and Perlman, 2015). Host proteases' cleavage of the spike glycoprotein is essential for virion entry into the target cells (Ou et al., 2020). The receptor-binding domain (RBD) in the S1 subunit binds ACE2 and facilitates fusion with host cell membrane (Tai et al., 2020). For the S2 domain of the spike glycoprotein, its function facilitates fusion of the viral and host cell membranes (Xia et al., 2020). Studying the SARS-CoV-2 *spike* (*S*) gene attracts a special attention, particularly from an immunologic and evolutionary points of view (Abdullahi et al., 2020; Chen et al., 2020; Korber et al., 2020; Robson, 2020). This gene is under an immune selective pressure, with neutralizing antibodies against its protein product inhibiting entry into the target cells (Korber et al., 2020; Walls et al., 2020). The swift genetic changes in the *S* gene can be used to infer the evolutionary relationships between viral sequences in a shorter time, compared to less variable regions (e.g. RNA-dependent RNA polymerase gene (*RdRp*)), where mutations occur, but appear to be more costly (Duffy et al., 2008; Moya et al., 2004; Pachetti et al., 2020; Robson, 2020). Genetic variability in the *S* gene can be demonstrated by continuous emergence of mutations, that were reported at a global level (Korber et al., 2020). Some of these mutations appeared to have a significant epidemiologic value, with the replacement of aspartic acid by glycine at position 614 of the spike glycoprotein (D614G), being associated with a higher viral shedding and increased infectivity (Korber et al., 2020; Maitra et al., 2020; Zhang et al., 2020). The increased infectivity can be attributed to increasing the probability of viral membrane fusion with target cell membrane through altering the receptor binding conformation (Volz et al., 2020; Yurkovetskiy et al., 2020). This mutation currently appears to be dominating the pandemic (Grubaugh et al., 2020). However, the clinical effect of such mutation is yet to be fully determined (Eaaswarkhanth et al., 2020; Kim et al., 2020b; Korber et al., 2020). Other mutations in the *S* gene have also been reported, with the most frequent including: D936Y/H, P1263L, and L5F (Korber et al., 2020; Lokman et al., 2020).

Similar to other RNA viruses, SARS-CoV-2 can be the subject of phylogenetic analysis due to its high evolutionary rate, and the application of molecular clock analysis might be of value to determine the timing of introductions of large clusters that imply networks of transmission (Duffy et al., 2008; Forster et al., 2020; Pybus and Rambaut, 2009). State-of-the-art methods for phylogeny construction include maximum likelihood and Bayesian tools (Anisimova et al., 2013).

The Middle East and North Africa (MENA) region was affected early on during the course of COVID-19 pandemic, with an overwhelming number of cases in some countries (e.g. Iran) (Karamouzian and Madani, 2020; Sawaya et al., 2020). The first confirmed cases of COVID-19 in the MENA dated back to February 2020 and were reported in UAE, Iran, and Egypt (Daw et al., 2020; Karamouzian and Madani, 2020; Mehtar et al., 2020). The total number of diagnosed cases of COVID-19 in the MENA exceeded 1,175,000 with more than 32,000 deaths reported as a result of the disease, as of July 25, 2020 (Worldometer, 2020).

Special attention to COVID-19 infections is needed in the countries of the MENA region, where political and economic factors might lead to devastating effects on the countries affected by the current pandemic (Karamouzian and Madani, 2020; Sawaya et al., 2020). Particular attention should be paid to countries like Yemen, Syria and Libya, where the ongoing instabilities can result in underreporting of COVID-19 cases and heavy burden on their health-care systems (Da'ar et al., 2020; Daw, 2020; Karamouzian and Madani, 2020; Sawaya et al., 2020).

The aims of this study included an attempt to phylogenetically analyze *S* gene sequences and to analyze the *S* gene mutation patterns in the MENA region. In addition, we aimed to characterize the temporal changes of D614G mutation spread in the region.

2. Materials and methods

2.1. Compilation of the MENA SARS-CoV-2 dataset

All SARS-CoV-2 sequences from the MENA countries (in the context of this work, MENA included the following 19 countries: Algeria, Bahrain, Egypt, Iran, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Palestine, Qatar, Kingdom of Saudi Arabia (KSA), Sudan, Syria, Tunisia, United Arab Emirates (UAE), and Yemen), were retrieved from the global science initiative and primary source for genomic data of influenza viruses (GISAID) (Elbe and Buckland-Merrett, 2017). We also downloaded the following sequence metadata if available: date of sequence collection, age, gender, city of collection together with country of sequence collection. The sequences were then aligned to the reference SARS-CoV-2 sequence (accession number: NC_045512) and alignment was conducted using multiple alignment program for amino acid or nucleotide sequences (MAFFT v.7) (Rozewicki et al., 2019). The MENA sequences that did not contain the complete *S* region were filtered out. In addition, we removed the sequences that contained indels, the nucleotide ambiguity (N); while other ambiguities were retained. The sequences that contained stop codons were removed as well. Each sequence header was also edited to include data in the following order: country of collection, collection date in days starting from January 5, 2020 (the date of reference sequence collection), city, accession number, gender, and age. The final dataset included 553 MENA *S* nucleotide sequences that were collected during January 2020 until June 2020.

2.2. Detection of the S gene mutations

Analysis of the full MENA SARS-CoV-2 *S* gene sequences was conducted in Molecular Evolutionary Genetics Analysis software (MEGA6) (Tamura et al., 2013). Visual inspection of the aligned MENA amino acid sequences was done, and mutations were identified based on comparison to the reference SARS-CoV-2 sequence (accession number: NC_045512), which was considered as the wild-type. Amino acids that were translated from codons containing ambiguous bases (e.g. R, Y), were excluded from mutation analysis.

2.3. Maximum likelihood phylogenetic analysis

The whole MENA *S* gene dataset was analyzed phylogenetically using the maximum likelihood (ML) approach in PhyML v3, with selection of the best nucleotide substitution model using Smart Model Selection (SMS), and depending on Akaike Information Criterion (AIC) (Guindon et al., 2010; Guindon and Gascuel, 2003; Lefort et al., 2017). The model which yielded the smallest AIC was the general time-reversible plus invariant sites (GTR + I) nucleotide substitution model with an estimated proportion of invariable sites of 0.625. The estimation of nodal support in the ML tree was based on the approximate Likelihood Ratio Test Shimodaira-Hasegawa like (aLRT-SH) with 0.90 as the statistical significance level (Anisimova et al., 2011). The ML analysis was repeated ten times and the ML tree with the highest likelihood was retained for final analysis, and determination of the MENA phylogenetic clusters was done by examining the ML tree from root to tips looking for branches with aLRT-SH $\geq$ 0.90, with large clusters having $\geq$15 sequences.

2.4. Bayesian estimation of time to most recent common ancestors (tMRCAs) of the large MENA phylogenetic clusters

For the large phylogenetic clusters (containing $\geq$15 sequences and identified using ML analysis), tMRCAs were estimated using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in BEAST v1.8.4 (Drummond et al., 2012). Bayesian analysis parameters included: Hasegawa–Kishono–Yano (HKY) nucleotide substitution model with discrete gamma-distributed rate heterogeneity, uncorrelated relaxed clock model with a normally-distributed rate prior (initial and mean

values of 0.0068, standard deviation = 0.0008), and a Bayesian skyline tree density model (Tang et al., 2020). For each large phylogenetic cluster, one run with 200 million chain length was performed. Samples of trees and parameters were collected every 20,000 steps after discarding a burn-in of 20%, and convergence was analyzed in Tracer v1.6.0 (Rambaut et al., 2015). The runs were accepted based on effective sample sizes (ESS) of ≥200 and convergence in the trace file. The maximum clade credibility (MCC) trees were assembled using TreeAnnotator in BEAST and were visualized using FigTree (Rambaut, 2012).

2.5. Statistical analysis

Chi-squared test (χ^2 test) was used to detect differences between the D614 and D614G groups in relation to gender and region (Middle East vs. North Africa). Mann-Whitney *U* test (M-W) was used to assess the difference between the D614 and D614G groups in relation to age. Linear-by-linear test for association (LBL) was used to assess the temporal changes in D614G prevalence. The statistical significance for all aforementioned tests was considered for $p < 0.050$.

2.6. Sequence accession numbers

A complete list of the MENA SARS-CoV-2 sequence epi accession numbers that were analyzed in this study is provided in (Appendix S1). These sequences are available publicly for registered users of GISAID (Shu and McCauley, 2017).

3. Results

3.1. The final MENA SARS-CoV-2 S gene sequence dataset

The total number of MENA SARS-CoV-2 *S* gene sequences that were included in final analysis was 553, distributed as follows: Oman (n = 159), KSA (n = 140), Egypt (n = 95), Morocco (n = 35), Bahrain (n = 34), UAE (n = 32), Jordan (n = 22), Tunisia (n = 8), Kuwait (n = 7), Qatar (n = 7), Lebanon (n = 6), Iran (n = 5), and Algeria (n = 3). The final length of the alignment was 3822 bases. Characteristics of the sequences are highlighted in (Table 1).

3.2. SARS-CoV-2 S gene mutations detected in the MENA

A total 55 unique non-synonymous mutations in the *S* gene were detected as compared to the reference SARS-CoV-2 genome. Eight

mutations were identified in spike receptor binding domain (SRD), compared to 21 mutations in S2 glycoprotein domain and 26 in other *S* regions. The most frequent mutation detected in the whole *S* region was D614G (n = 435), followed by Q677H (n = 8), and V6F (n = 5). The majority of mutations were detected sporadically (n = 43, 78.2%, Table 2). The highest number of unique *S* gene mutations (including D614G) was noticed in Oman (n = 16), followed by Egypt (n = 15), Bahrain (n = 9), and KSA (n = 6, Table 2).

3.3. Variables associated with a higher prevalence of D614G mutation

Analysis of the two variants of *S* gene (D614 vs. D614G) showed a higher prevalence of D614G in North Africa compared to the Middle East (95.0% vs. 73.7%, $p < 0.001$; χ^2 test). In addition, a higher prevalence of D614G variant was noticed in the second half of the study period (April, May and June vs. January, February and March, 90.7% vs. 59.5%, $p < 0.001$; χ^2 test). However, no statistical difference was noticed upon comparing the two variants based on age ($p = 0.195$; M-W), age group (less than 40 years vs. more than or equal to 40 years, $p = 0.176$; χ^2 test), or gender ($p = 0.644$; χ^2 test). Analysis of the D614G mutant per country showed its presence in all MENA countries included in the study with exception of Iran and Qatar (Figure 1). In addition, no statistical difference was found in analysis per country upon comparing the two variants based on age, age group, or gender.

3.4. Temporal trend of D614G mutant spread in the MENA

Analysis of temporal trend of spread of the D614G mutant of SARS-CoV-2 in the whole MENA region as a single unit revealed an increasing prevalence of D614G from 63.0% in January 2020 to reach 98.5% in June 2020 ($p < 0.001$; LBL, Figure 2). The same pattern was detected upon comparing the first three months of 2020, compared to April, May and June 2020 (59.5% vs. 90.7%; $p < 0.001$; χ^2 test).

3.5. Maximum likelihood phyloegentic tree of MENA S gene sequences

To assess the possible presence of phylogenetic clusters in the MENA, ML analysis was conducted. The constructed ML tree showed a star-shaped pattern with short internal branches and long terminal branches (Appendix S2). A total of 13 phylogenetic clusters (aLRT-SH ≥ 0.9) were determined; eight of which included sequences from a single MENA country and five clusters contained sequences collected in more than one MENA country (Appendix S2). Five clusters contained two

Table 1. Characteristics of SARS-CoV-2 sequences collected in the Middle East and North Africa and its metadata.

Country	Number of sequences	Age (mean, SD[3])	Gender N[4] (%)		Period for sequence collection
			Male	Female	
Oman	159	38 (16.8)	82 (51.9)	76 (48.1)	23-02-2020 to 11-06-2020
KSA[1]	140	42 (16.6)	68 (74.7)	23 (25.3)	03-02-2020 to 20-04-2020
Egypt	95	41 (14.4)	20 (60.6)	13 (39.4)	18-03-2020 to 20-06-2020
Morocco	35	36 (6.6)	7 (100.0)	0	27-02-2020 to 21-05-2020
Bahrain	34	-	-	-	07-03-2020 to 25-06-2020
UAE[2]	32	37 (13.8)	20 (64.5)	11 (35.5)	29-01-2020 to 04-05-2020
Jordan	22	-	-	-	16-03-2020 to 08-04-2020
Tunisia	8	-	1 (50.0)	1 (50.0)	18-03-2020 to 10-04-2020
Kuwait	7	-	2 (100.0)	0	02-03-2020 to 16-03-2020
Qatar	7	-	-	-	23-03-2020
Lebanon	6	49 (17.1)	3 (50.0)	3 (50.0)	27-02-2020 to 15-03-2020
Iran	5	-	-	-	09-03-2020 to 29-03-2020
Algeria	3	-	-	-	02-03-2020 to 08-03-2020

[1] KSA: Kingdom of Saudi Arabia.
[2] UAE: United Arab Emirates.
[3] SD: Standard deviation.
[4] N: Number. Notice that results for age were not mentioned if the number of available sequences were less than 5.

Table 2. Amino acid substitutions in the spike (S) protein of SARS-CoV-2 that were detected in the Middle East and North Africa (MENA), stratified by domain.

Spike protein region	Mutation (country, number of sequences that contained the mutation)
Spike receptor binding domain	**R408I** (Egypt, n[2] = 2), **A570S** (Egypt, n = 1), **A522V** (Egypt, n = 1), **S514Y** (Oman, n = 1), **P499H** (Egypt, n = 1), **S477R** (Egypt, n = 1), **S459F** (Bahrain, n = 1), **A344S** (Saudi Arabia, n = 1)
S2 glycoprotein	**Q677H** (Egypt, n = 8), **H1101Y** (Oman, n = 4), **A958S** (Saudi Arabia, n = 2), **C1243F** (Oman, n = 1), **M1237I** (Morocco, n = 1), **V1228L** (Oman, n = 1), **V1176F** (Egypt, n = 1), **A1174V** (Oman, n = 1), **G1167S** (Jordan, n = 1), **D1153A** (Egypt, n = 2), **D1146Y** (Oman, n = 2), **D1139Y** (Jordan, n = 1), **L1063F** (Bahrain, n = 1), **S939F** (UAE[3], n = 1), **D936Y** (Oman, n = 1), **A871S** (Bahrain, n = 1), **T859I** (Oman, n = 1), **I850F** (UAE, n = 1), **T732S** (Egypt, n = 1), **M731I** (Saudi Arabia, n = 1), **A684V** (Saudi Arabia, n = 1)
Others[1]	**D614G[4]** (n = 435), **V6F** (Morocco, n = 5), **L5F** (Oman, n = 2; Egypt, n = 1; Morocco, n = 1), **S640 A/F** (Egypt, n = 1; Oman, n = 1), **V622I/F** (Bahrain, n = 1, Oman = 1), **M177I** (Bahrain, n = 2), **A653V** (Egypt, n = 1), **P621S** (KSA, n = 1), **Q314R** (Tunisia, n = 1), **G311E** (Morocco, n = 1), **A288T** (Tunisia, n = 1), **Y279N** (Tunisia, n = 1), **A263V** (UAE, n = 1), **A262T** (Oman, n = 1), **S255F** (Bahrain, n = 1), **M153I** (Lebanon, n = 1), **P138H** (Egypt, n = 1), **T95I** (Oman, n = 1), **G75S** (Bahrain, n = 1), **A67S** (Bahrain, n = 1), **T29I** (Tunisia, n = 1), **Y28H** (UAE, n = 1), **T22I** (Iran, n = 1), **R21I** (Oman, n = 1), **S13I** (Oman, n = 1), **S12F** (Egypt, n = 1)

[1] Others: Amino acid substitutions in regions other than the spike receptor binding domain and S2 glycoprotein.
[2] n: Number.
[3] UAE: United Arab Emirates
[4] D614G: The replacement of aspartic acid by glycine at position 614 of the spike glycoprotein, which dominated the sequences and that were analyzed separately in the main manuscript.

sequences, and two large clusters were identified, each containing 26 MENA sequences. The highest percentage of clustering sequences was found in Iran (n = 3/5, 60.0%), followed by KSA (n = 39/149, 27.9%), and Tunisia (n = 2/8, 25.0%, Figure 3). The overall proportion of phylogenetic clustering was 15.4% (n = 85/553).

3.6. Bayesian analysis of the largest MENA phylogenetic clusters

Bayesian phylogenetic analysis was conducted on the two large clusters identified previously using the ML approach. One Egyptian sequence was removed from each cluster due to the lack of exact collection date (EPI_ISL_475753 for the first cluster and EPI_ISL_475746 for the second cluster). This resulted in analysis of two clusters, each containing 25 sequences. The first cluster contained 14 Saudi sequences, ten Omani sequences and a single Egyptian sequence, with a range of sequence collection between February 13 and May 11. The median estimate for tMRCA for this cluster having the D614G mutation was February 8, 2020 (95% highest posterior density interval [HPD]: October

19, 2019–February 13, 2020, Figure 4). For the second cluster (D614) with 20 Saudi sequences, three Egyptian sequences and two Tunisian sequences, the estimated median tMRCA was March 15, 2020 (95% HPD: February 21, 2020–March 15, 2020). The mean evolutionary rate estimated by molecular clock analysis was 6.46×10^{-3} substitutions/site/year (s/s/y) for the first cluster (95% HPD: 4.87×10^{-3} - 8.03×10^{-3} s/s/y), and 6.50×10^{-3} s/s/y for the second cluster (95% HPD: 4.91×10^{-3} - 8.03×10^{-3} s/s/y).

4. Discussion

In this study, phylogenetic analysis tools were utilized to assess origins, spread and mutations of SARS-CoV-2 in the MENA. Phylogeny construction can help to formulate hypotheses regarding the spread of certain taxa having a common origin (Ciccozzi et al., 2019; Pybus and Rambaut, 2009). In addition, molecular clock analysis can help to establish a timeline for origins of monophyletic clades (Jenkins et al., 2002; Nasir and Caetano-Anolles, 2015). Phylogenetic analysis of the
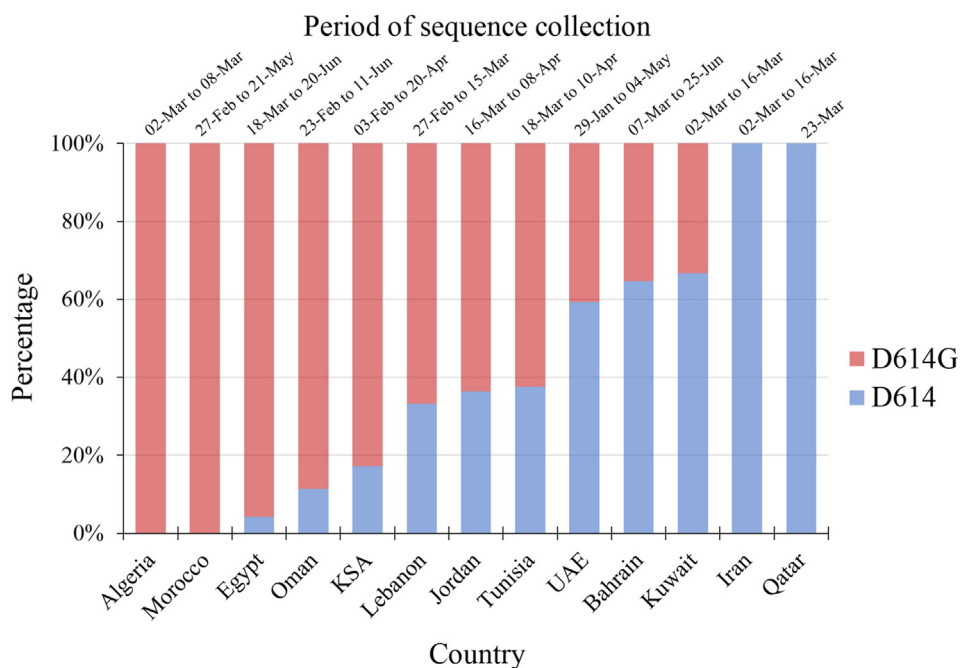

Figure 1. The relative proportions of D614 and D614G mutation in the Middle East and North Africa stratified by countries of SARS-CoV-2 sequence collection. KSA: Kingdom of Saudi Arabia, UAE: United Arab Emirates, SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2. The period of sequence collection varies depending on the country, which is shown in the upper part of the figure.
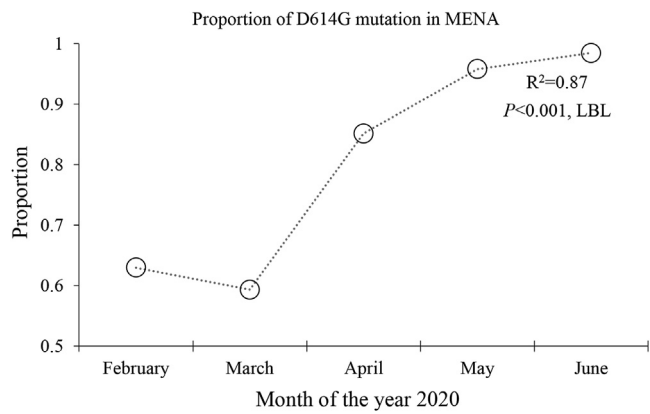
Proportion of D614G mutation in MENA

Figure 2. Temporal change in the prevalence of D614G in the Middle East and North Africa stratified by months of SARS-CoV-2 sequence collection. SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2, LBL: Linear-by-linear test for association.

MENA *S* gene SARS-CoV-2 sequences showed a relatively low level of phylogenetic clustering (15%), which hints to a large number of virus introductions into the region. In addition, molecular clock analysis suggests an early introduction of the virus into the MENA which might have been circulating in the region from early February 2020 or even earlier, with subsequent spread into large networks of virus transmission. This estimate of an early virus introduction is supported by the close proximity in time of official reporting of confirmed COVID-19 cases in the region (Karamouzian and Madani, 2020). Moreover, an evidence of inter-country spread of the virus was manifested by the presence of mixing between Middle Eastern and North African taxa in the two large MENA clusters, which hints to an early spread of the virus among the countries of the region.

In this study, no evidence of distinct SARS-CoV-2 genetic variants was found. Plausible explanations might be related to the use of sub-genomic part of the genome (the *S* gene) rather than utilizing the whole genome, similar to the previous study by Yang et al., (2020). The rationale behind selecting the *S* region was for two reasons: first, the variability of this region is expected to be higher than other parts of the genome (e.g. *RdRp*, where mutations are more costly) (Agostini et al., 2018; Shannon et al.,

2020). Second, mutations in the *S* gene can have significant impact particularly for vaccine development and utility of neutralizing antibodies (Lokman et al., 2020). The absence of distinct SARS-CoV-2 genetic variants in this study does not provide a conclusive evidence of its genuine absence from the region. These two genetic variants (named L and S lineages) were reported previously, however, a recent report by MacLean et al. carefully discussed the potential pitfalls of such premature conclusions (MacLean et al., 2020; Tang et al., 2020).

For the estimated evolutionary rate of the two large MENA clusters identified in this study, we based the rate prior selection on the previous finding by (Giovanetti et al., 2020). This estimate appears higher than other estimates for SARS-CoV-2 and should be interpreted with caution based on our selection of a strong prior. However, the rate estimate might appear plausible, since it represents the *S* gene, rather than the whole genome. For ML analysis, the MENA sequences yielded a star-like phylogeny suggesting a recent growing epidemic (Colijn and Plazzotta, 2018).

The major result of this study was the demonstration of a temporal shift of SARS-CoV-2 from D614 into D614G variant, which dominated the most recent sequences collected in the region. Such trend was revealed at the global level by Korber et al., and our results indicated a similar pattern in the MENA (Korber et al., 2020). In the aforementioned comprehensive study, Korber et al. estimated the global prevalence of D614G at 71.0%, whereas our estimate in the MENA was 78.7%, which appears reasonable, bearing in mind the protracted duration of sequence collection in this study. The explanation for such an observation is most likely related to the association of D614G with a higher viral load and subsequent higher quantities of the virus shed by infected individuals, which increases the likelihood of infection by such a mutant, although an early founder effect of this variant cannot be ruled out (Deng et al., 2020; Farkas et al., 2020; Yurkovetskiy et al., 2020; Zhang et al., 2020). Whether this variant can have an effect on severity and outcome of COVID-19 is yet to be fully determined (Becerra-Flores and Cardozo, 2020; Eaaswarkhanth et al., 2020; Korber et al., 2020). This mutation appeared in all MENA countries, except in Qatar and Iran, which might be related to the low number of sequences from these two countries that were found in GISAID, and the early time of sequence collection (less than 10 sequences from each country were found, dating back to March, 2020). The emergence of D614G and its increasing prevalence have been reported by several published papers and preprints including a report from North Africa by Laamarti et al., albeit with a fewer number of
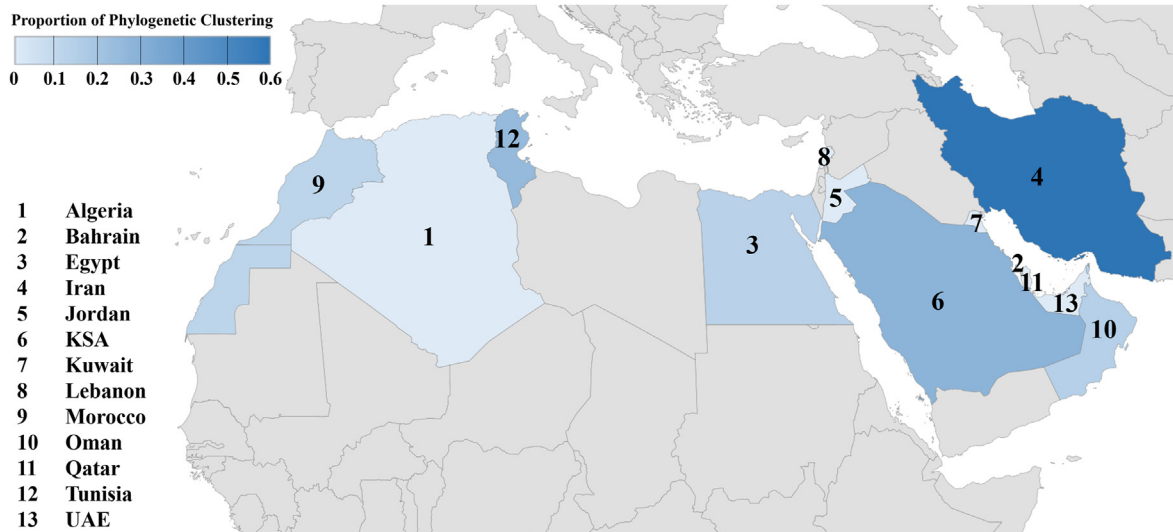

Figure 3. The Middle East and North Africa (MENA) map showing the proportion of phylogenetic clustering among the spike (*S*) sequences as inferred by maximum likelihood phylogenetic analysis. The upper left legend indicates the proportion of shown by different shades of blue. The country names were replaced by numbers on the map to increase the visibility. SA: Kingdom of Saudi Arabia, UAE: United Arab Emirates. Other MENA countries that lacked sequences are not shown in the blue scale. The figure was generated in Microsoft Excel, powered by Bing, © GeoNames, Microsoft, Navinfo, TomTom, Wikipedia.
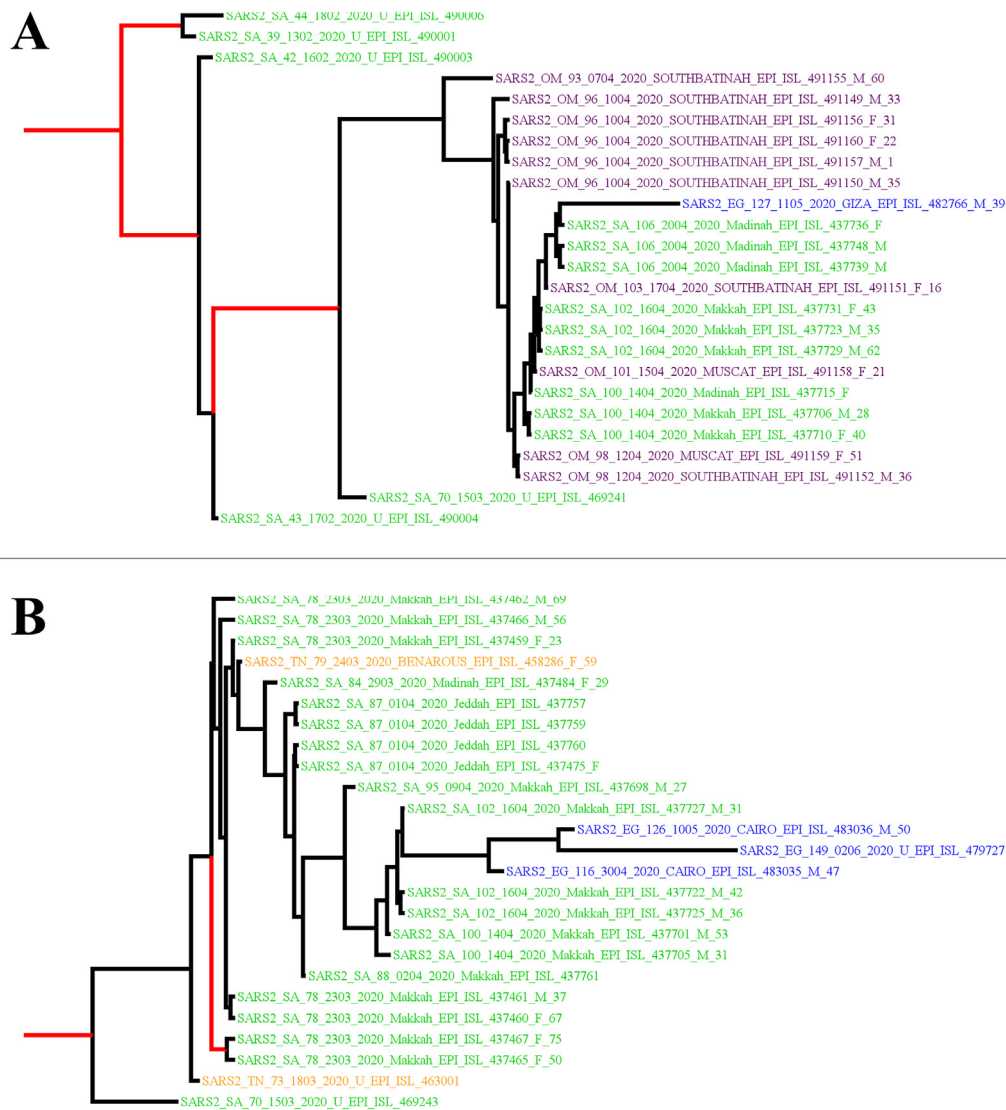
Figure 4. Maximum clade credibility (MCC) trees of the two large Middle East and North Africa (MENA) SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) phylogenetic clusters. A) The upper MCC tree with sequences having the D614G mutation. B) The lower MCC tree represents the D614 cluster. The sequence names were colored based on country of collection (Egypt [EG]: blue, Oman [OM]: purple, Saudi Arabia [SA]: green, and Tunisia [TN]: orange). The sequences were named based on the following: Country of sequence collection, day, month and year of sequence collection and SARS-CoV-2 sequence epi accession numbers. Internal branches with posterior values ≥0.70 are shown in red.

sequences than the one analyzed in the current study (Gong et al., 2020; Kim et al., 2020b; Laamarti et al., 2020; Maitra et al., 2020).

Other amino acid replacements that were found in the study included Q677H (found only in Egypt), and L5F found in three different countries (Oman, Egypt, and Morocco). The L5F replacement is located in the signal peptide domain of the spike glycoprotein and might be related to recurring sequencing errors (Korber et al., 2020; De Maio et al., 2020). Nevertheless, its appearance in different studies warrants further investigation to determine its significance (Korber et al., 2020). The functional importance of Q677H replacement as not been determined yet despite a previous report describing its occurrence (Kim et al., 2020a).

Limitations of this study should be clearly stated and taken into consideration. The most obvious caveat in the study was sampling bias, in time and location. This was particularly reflected in the predominance of Omani and Saudi sequences in the large clusters. In spite of reporting COVID-19 in all MENA countries, the following countries did not have *S* sequences submitted to GISAID: Syria, Libya, Yemen, Sudan and Palestine (Iraq had partial sequences that did not include the *S* gene). In addition, bias was observed for timing of

sequence collection. Furthermore, only two countries (Oman and KSA) had more than 100 sequences available for analysis. Another point that should be considered is related to the molecular clock analysis, where we used a strong informative prior which may have affected our tMRCA estimates for dating the origins of the two large phylogenetic clusters. Sequencing errors should also be taken into account, which can partly explain some sporadic mutations that were found in this study.

5. Conclusions

In the current study, we demonstrated that the D614G variant of SARS-CoV-2 appears to be taking over COVID-19 epidemic in the MENA, similar to what have been reported in other regions around the globe. Local transmission of SARS-CoV-2 might have been established earlier than previously thought, and this illustrates the importance of vigilant surveillance in such conditions of outbreaks by novel viruses. The mutational patterns of SARS-CoV-2 should be closely monitored as the virus seems to be heading into an endemicity in the human population,

particularly in relation to mutations' potential impact on passive and active immunization.

Declarations

Author contribution statement

Malik Sallam: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Nidaa A. Ababneh, Deema Dababseh, Faris G. Bakri: Analyzed and interpreted the data; Wrote the paper.

Azmi Mahafzah: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The datasets analysed during the current study are available from the corresponding author on reasonable request and considering the terms of use by GISAID.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2021.e06035.

References

Abdullahi, I.N., Emeribe, A.U., Ajayi, O.A., Oderinde, B.S., Amadu, D.O., Osuji, A.I., 2020. Implications of SARS-CoV-2 genetic diversity and mutations on pathogenicity of COVID-19 and biomedical interventions. J. Taibah Univ. Med. Sci.

Agostini, M.L., Andres, E.L., Sims, A.C., Graham, R.L., Sheahan, T.P., Lu, X., Smith, E.C., Case, J.B., Feng, J.Y., Jordan, R., Ray, A.S., Cihlar, T., Siegel, D., Mackman, R.L., Clarke, M.O., Baric, R.S., Denison, M.R., 2018. Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. mBio 9.

Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., Gascuel, O., 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60, 685–699.

Anisimova, M., Liberles, D.A., Philippe, H., Provan, J., Pupko, T., von Haeseler, A., 2013. State-of the art methodologies dictate new standards for phylogenetic analysis. BMC Evol. Biol. 13, 161.

Becerra-Flores, M., Cardozo, T., 2020. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. Int. J. Clin. Pract., e13525

Chen, W.H., Hotez, P.J., Bottazzi, M.E., 2020. Potential for developing a SARS-CoV receptor-binding domain (RBD) recombinant protein as a heterologous human vaccine against coronavirus infectious disease (COVID)-19. Hum. Vaccines Immunother. 16, 1239–1242.

Cherry, J.D., Krogstad, P., 2004. SARS: the first pandemic of the 21st century. Pediatr. Res. 56, 1–5. FC.

Ciccozzi, M., Lai, A., Zehender, G., Borsetti, A., Cella, E., Ciotti, M., Sagnelli, E., Sagnelli, C., Angeletti, S., 2019. The phylogenetic approach for viral infectious disease evolution and epidemiology: an updating review. J. Med. Virol. 91, 1707–1724.

Colijn, C., Plazzotta, G., 2018. A metric on phylogenetic tree shapes. Syst. Biol. 67, 113–126.

Da'ar, O.B., Haji, M., Jradi, H., 2020. Coronavirus Disease 2019 (COVID-19): potential implications for weak health systems and conflict zones in the Middle East and North Africa region. Int. J. Health Plann. Manag. 1–6.

Daw, M., El-Bouzedi, A., Ahmed, M., Cheikh, Y., 2020. Spatial distri-bution and geographic mapping of COVID-19 in Northern African countries; A preliminary study. J. Clin. Immunol. Immunother. 6, 32.

Daw, M.A., 2020. Corona virus infection in Syria, Libya and Yemen; an alarming devastating threat. Trav. Med. Infect. Dis. 101652.

Delabouglise, A., Choisy, M., Phan, T.D., Antoine-Moussiaux, N., Peyre, M., Vu, T.D., Pfeiffer, D.U., Fournie, G., 2017. Economic factors influencing zoonotic disease dynamics: demand for poultry meat and seasonal transmission of avian influenza in Vietnam. Sci. Rep. 7, 5905.

Deng, X., Gu, W., Federman, S., du Plessis, L., Pybus, O.G., Faria, N., Wang, C., Yu, G., Bushnell, B., Pan, C.Y., Guevara, H., Sotomayor-Gonzalez, A., Zorn, K., Gopez, A., Servellita, V., Hsu, E., Miller, S., Bedford, T., Greninger, A.L., Roychoudhury, P., Starita, L.M., Famulare, M., Chu, H.Y., Shendure, J., Jerome, K.R., Anderson, C., Gangavarapu, K., Zeller, M., Spencer, E., Andersen, K.G., MacCannell, D., Paden, C.R., Li, Y., Zhang, J., Tong, S., Armstrong, G., Morrow, S., Willis, M., Matyas, B.T., Mase, S., Kasirye, O., Park, M., Masinde, G., Chan, C., Yu, A.T., Chai, S.J., Villarino, E., Bonin, B., Wadford, D.A., Chiu, C.Y., 2020. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. Science 369, 582–587.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Duffy, S., Shackelton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. 9, 267–276.

Eaaswarkhanth, M., Al Madhoun, A., Al-Mulla, F., 2020. Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? Int. J. Infect. Dis. 96, 459–460.

Elbe, S., Buckland-Merrett, G., 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob. Chall. 1, 33–46.

Farkas, C., Fuentes-Villalobos, F., Garrido, J.L., Haigh, J., Barria, M.I., 2020. Insights on early mutational events in SARS-CoV-2 virus reveal founder effects across geographical regions. PeerJ 8, e9255.

Fehr, A.R., Perlman, S., 2015. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol. Biol. 1282, 1–23.

Forster, P., Forster, L., Renfrew, C., Forster, M., 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc. Natl. Acad. Sci. U. S. A. 117, 9241–9243.

Giovanetti, M., Benvenuto, D., Angeletti, S., Ciccozzi, M., 2020. The first two cases of 2019-nCoV in Italy: where they come from? J. Med. Virol. 92, 518–521.

Gong, Y.N., Tsao, K.C., Hsiao, M.J., Huang, C.G., Huang, P.N., Huang, P.W., Lee, K.M., Liu, Y.C., Yang, S.L., Kuo, R.L., Chen, K.F., Liu, Y.C., Huang, S.Y., Huang, H.I., Liu, M.T., Yang, J.R., Chiu, C.H., Yang, C.T., Chen, G.W., Shih, S.R., 2020. SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. Emerg. Microb. Infect. 9, 1457–1466.

Graham, R.L., Baric, R.S., 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. 84, 3134–3146.

Grubaugh, N.D., Hanage, W.P., Rasmussen, A.L., 2020. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. Cell.

Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., Butt, K.M., Wong, K.L., Chan, K.W., Lim, W., Shortridge, K.F., Yuen, K.Y., Peiris, J.S., Poon, L.L., 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302, 276–278.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. 54, 156–165.

Karamouzian, M., Madani, N., 2020. COVID-19 response in the Middle East and north Africa: challenges and paths forward. Lancet Glob. Health 8, e886–e887.

Karesh, W.B., Dobson, A., Lloyd-Smith, J.O., Lubroth, J., Dixon, M.A., Bennett, M., Aldrich, S., Harrington, T., Formenty, P., Loh, E.H., Machalaba, C.C., Thomas, M.J., Heymann, D.L., 2012. Ecology of zoonoses: natural and unnatural histories. Lancet 380, 1936–1945.

Kim, J.S., Jang, J.H., Kim, J.M., Chung, Y.S., Yoo, C.K., Han, M.G., 2020a. Genome-wide identification and characterization of point mutations in the SARS-CoV-2 genome. Osong Publ. Health Res. Perspect. 11, 101–111.

Kim, S.J., Nguyen, V.G., Park, Y.H., Park, B.K., Chung, H.C., 2020b. A novel synonymous mutation of SARS-CoV-2: is this possible to affect their antigenicity and immunogenicity? Vaccines (Basel) 8.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Sheffield, C.-G.G., McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182, 812–827.

Laamarti, M., Kartti, S., Alouane, T., Laamarti, R., Allam, L., Ouadghiri, M., Chemao-Elfihri, M.W., Smyej, I., Rahoui, J., Benrahma, H., Diawara, I., Essabbar, A., Boumajdi, N., Bendani, H., Bouricha, E.M., Aanniz, T., Elattar, J., Hafidi, N.E., Jaoudi, R.E., Sbabou, L., Nejjari, C., Amzazi, S., Mentag, R., Belyamani, L., Ibrahimi, A., 2020. Genetic analysis of SARS-CoV-2 strains collected from North Africa: Viral Origins and Mutational Spectrum. bioRxiv 2020.

Lefort, V., Longueville, J.E., Gascuel, O., 2017. SMS: Smart model selection in PhyML. Mol. Biol. Evol. 34, 2422–2424.

Liu, Y.C., Kuo, R.L., Shih, S.R., 2020. COVID-19: the first documented coronavirus pandemic in history. Biomed. J.

Lokman, S.M., Rasheduzzaman, M., Salauddin, A., Barua, R., Tanzina, A.Y., Rumi, M.H., Hossain, M.I., Siddiki, A., Mannan, A., Hasan, M.M., 2020. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach. Infect. Genet. Evol. 84, 104389.

Lu, G., Liu, D., 2012. SARS-like virus in the Middle East: a truly bat-related coronavirus causing human diseases. Protein Cell 3, 803–805.

MacLean, O.A., Orton, R.J., Singer, J.B., Robertson, D.L., 2020. No evidence for distinct types in the evolution of SARS-CoV-2. Virus Evol. 6.

Maitra, A., Sarkar, M.C., Raheja, H., Biswas, N.K., Chakraborti, S., Singh, A.K., Ghosh, S., Sarkar, S., Patra, S., Mondal, R.K., Ghosh, T., Chatterjee, A., Banu, H., Majumdar, A., Chinnaswamy, S., Srinivasan, N., Dutta, S., Das, S., 2020. 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.

Mehtar, S., Preiser, W., Lakhe, N.A., Bousso, A., TamFum, J.M., Kallay, O., Seydi, M., Zumla, A., Nachega, J.B., 2020. Limiting the spread of COVID-19 in Africa: one size mitigation strategies do not fit all countries. Lancet Glob. Health 8, e881–e883.

Morse, S.S., Mazet, J.A., Woolhouse, M., Parrish, C.R., Carroll, D., Karesh, W.B., Zambrana-Torrelio, C., Lipkin, W.I., Daszak, P., 2012. Prediction and prevention of the next pandemic zoonosis. Lancet 380, 1956–1965.

Moya, A., Holmes, E.C., Gonzalez-Candelas, F., 2004. The population genetics and evolutionary epidemiology of RNA viruses. Nat. Rev. Microbiol. 2, 279–288.

De Maio, N., CW, Borges, R., Weilguny, L., Slodkowicz, G., Goldman, N., 2020. Issues with SARS-CoV-2 sequencing data. Available at. https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473. (Accessed 25 July 2020).

Nasir, A., Caetano-Anolles, G., 2015. A phylogenomic data-driven exploration of viral origins and evolution. Sci. Adv. 1, e1500527.

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., Jin, Q., Wang, J., Qian, Z., 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat. Commun. 11, 1620.

Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., Masciovecchio, C., Angeletti, S., Ciccozzi, M., Gallo, R.C., Zella, D., Ippodrino, R., 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J. Transl. Med. 18, 179.

Peiris, J.S., Yuen, K.Y., Osterhaus, A.D., Stohr, K., 2003. The severe acute respiratory syndrome. N. Engl. J. Med. 349, 2431–2441.

Pybus, O.G., Rambaut, A., 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nat. Rev. Genet. 10, 540–550.

Rambaut, A., 2012. FigTree V1. 4. Available at. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 2017 2017.

Rambaut, A., Suchard, M., Xie, D., Drummond, A., 2015. Tracer v1. 6.

Robson, B., 2020. 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. 121, 103749.

Rozewicki, J., Li, S., Amada, K.M., Standley, D.M., Katoh, K., 2019. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 47, W5–W10.

Sawaya, T., Ballouz, T., Zaraket, H., Rizk, N., 2020. Coronavirus disease (COVID-19) in the Middle East: a call for a unified response. Front. Publ. Health 8, 209.

Sevajol, M., Subissi, L., Decroly, E., Canard, B., Imbert, I., 2014. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. Virus Res. 194, 90–99.

Shannon, A., Le, N.T., Selisko, B., Eydoux, C., Alvarez, K., Guillemot, J.C., Decroly, E., Peersen, O., Ferron, F., Canard, B., 2020. Remdesivir and SARS-CoV-2: structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. Antivir. Res. 178, 104793.

Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 22.

Su, S., Wong, G., Shi, W., Liu, J., Lai, A.C.K., Zhou, J., Liu, W., Bi, Y., Gao, G.F., 2016. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol. 24, 490–502.

Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., Zhou, Y., Du, L., 2020. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cell. Mol. Immunol. 17, 613–620.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., Lu, J., 2020. On the origin and continuing evolution of SARS-CoV-2. Nat. Sci. Rev. 7, 1012–1023.

Volz, E., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., O'Toole, A., Southgate, J., Johnson, R., Jackson, B., Nascimento, F.F., Rey, S.M., Nicholls, S.M., Colquhoun, R.M., da Silva Filipe, A., Shepherd, J., Pascall, D.J., Shah, R., Jesudason, N., Li, K., Jarrett, R., Pacchiarini, N., Bull, M., Geidelberg, L., Siveroni, I., Consortium, C.-U., Goodfellow, I., Loman, N.J., Pybus, O.G., Robertson, D.L., Thomson, E.C., Rambaut, A., Connor, T.R., 2020. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 181, 281–292 e286.

Woo, P.C., Lau, S.K., Huang, Y., Yuen, K.Y., 2009. Coronavirus diversity, phylogeny and interspecies jumping. Exp. Biol. Med. 234, 1117–1127.

Worldometer, 2020. COVID-19 Coronavirus pandemic. Available at. https://www.worldometers.info/coronavirus/. (Accessed 25 July 2020).

Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., Qi, F., Bao, L., Du, L., Liu, S., Qin, C., Sun, F., Shi, Z., Zhu, Y., Jiang, S., Lu, L., 2020. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. Cell Res. 30, 343–355.

Yang, X., Dong, N., Chan, E.W.-C., Chen, S., 2020. Genetic cluster analysis of SARS-CoV-2 and the identification of those responsible for the major outbreaks in various countries. Emerg. Microb. Infect. 9, 1287–1299.

Yurkovetskiy, L., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T., Wang, Y., Baum, A., Diehl, W.E., Dauphin, A., Carbone, C., Veinotte, K., Egri, S.B., Schaffner, S.F., Lemieux, J.E., Munro, J., Sabeti, P.C., Kyratsous, C.A., Shen, K., Luban, J., 2020. SARS-CoV-2 Spike protein variant D614G increases infectivity and retains sensitivity to antibodies that target the receptor binding domain. bioRxiv 2020.

Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Rangarajan, E.S., Izard, T., Farzan, M., Choe, H., 2020. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv.