Title: No association between the SARS-CoV-2 variants and mortality rates in the Eastern Mediterranean Region

Authors: Saad Omais[1], Samer Kharroubi[2], Hassan Zaraket[3,4,#]

Affiliations:

1, Department of Biology, Faculty of Arts and Sciences, American University of Beirut, Beirut, Lebanon.

2, Department of Nutrition and Food Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut, Beirut, Lebanon.

3, Faculty of Medicine, Center for Infectious Diseases Research, American University of Beirut, 1107 2020 Beirut, Lebanon.

4, Department of Experimental Pathology, Faculty of Medicine, Immunology & Microbiology, American University of Beirut, 1107 2020 Beirut, Lebanon.

 # Corresponding author: hz34@aub.edu.lb

 # Present/permanent address: Department of Experimental Pathology, Faculty of Medicine, Immunology & Microbiology, American University of Beirut, 1107 2020 Beirut, Lebanon

Keywords: SARS-CoV-2, whole genome, variants, case fatality rate, East Mediterranean Region

Highlights:

1. Analysis of SARS-CoV-2 whole genome sequences from Eastern Mediterranean Region
2. Large data gaps exist in the region
3. Spike_D614G and NSP12_P323L, were predominant in most countries
4. No association between common variants and case fatality rates

Abstract

As the novel coronavirus SARS-CoV-2 continues to spread in all countries, there is a growing interest in monitoring and understanding the impact of emerging strains on virus transmission and disease severity. Here, we analyzed SARS-CoV-2 genomic sequences collected in the Eastern Mediterranean Region (EMR) countries which, as of 12 November 2020. The majority (73.7%) of these sequences originated from four out of 22 EMR countries, and 65.6% of all sequences belonged to GISAID clades GR, GH, and G. We identified ten common non-synonymous mutations represented among SARS-CoV-2 in the EMR and several country-specific ones. Two substitutions, S_D614G and NSP12_P323L, were predominant in most countries. Contrary to previous studies, these substitutions were not consistently associated with increased case fatality rates across the Region. Our study identified critical data gaps in EMR highlighting the importance of enhancing surveillance and sequencing capacities in the region.

Introduction

In early October 2020, the World Health Organization (WHO) estimated that 10% of the global population has been infected with the novel coronavirus, SARS-CoV-2 [1]. Since its declaration as a pandemic seven months earlier, COVID-19 has inflicted substantial health and economic burden around the world [2, 3]. As of 21 November 2020, there have been around 56 million confirmed cases and 1.36 million deaths attributed to COVID-19 [4]. Specifically, 3,763,231 confirmed cases (of whom 3,096,119 have recovered) and 95,558 deaths were reported in the Eastern Mediterranean Region (EMR) [5].

Prior to SARS-CoV-2 emergence, six human coronavirus (HCoV) strains were known to infect humans: two of which caused fatal respiratory diseases (SARS-CoV-1 and MERS-CoV), while the remaining four circulate annually and cause 15-29% of all common colds (HCoV-229E, -OC43, -NL63 and -HKU1) [6]. SARS-CoV-2 gains access to the host cell by binding to the human angiotensin-converting enzyme 2 (hACE2) through its spike (S) protein. SARS-CoV-2 S receptor-binding domain (RBD) has a 10- to 20-fold higher hACE2 binding affinity than SARS-CoV-1 [7]. Yet, SARS-CoV-2 RBD is less accessible such that the overall S-hACE2 interaction is similar or weaker than that of SARS-CoV-1. Nonetheless, the SARS-CoV-2 spike protein possesses a polybasic cleavage site making it accessible to furin protease which enhances cell entry [8]. Similar to other beta-coronaviruses, SARS-CoV-2 has a positive-sense RNA genome that is around 30-kb in length with six functional open reading frames (ORFs). ORFa and -b constitute nearly two-thirds of the genome and produce 16 non-structural proteins (NSPs), including the RNA-dependent RNA polymerase (RdRp or NSP12) and a helicase (NSP13). The structural nucleocapsid (N), membrane (M), envelope (E), and S proteins are encoded by the remaining stretch of the RNA genome along with other accessory proteins [9].

There has been a tremendous global surveillance effort to closely monitor SARS-CoV-2 circulating worldwide that enabled timely detection of emerging variants. SARS-CoV-2 evolves at an estimated rate of around 6 x 10^{-4} nucleotides/genome/year [10]. Numerous variants have emerged but only a handful of them have been fixated such as the S_D614G that has prevailed globally. S_D614G was shown to alter infectivity and virulence of SARS-CoV-2 and was associated with an increase in mortality [10-14]. Recently, Sallam et al. reported in a preprint an increase in the incidence of S_D614G variant between February and June 2020 in the Middle East and North Africa [15]. However, a detailed genomic characterization of SARS-CoV-2 in EMR is still lacking. It also remains not clear whether the case fatality rates (CFRs) in EMR are associated with any of the prevailing SARS-CoV-2 variants. In this study, we analyzed SARS-CoV-2 whole-genome sequences collected in EMR and examined the temporal and country-level associations between the predominant variants and CFRs.

Methods

Sequence Retrieval and Analysis

SARS-CoV-2 sequences were downloaded from GISAID EpiCoV™ database [16] which were submitted by 12 November 2020 for EMR countries. These are Afghanistan, Bahrain, Djibouti, Egypt, Iran, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Pakistan, Palestine, Qatar, Saudi Arabia, Somalia, Sudan, Syrian Arab Republic, Tunisia, United Arab Emirates (UAE), and Yemen (http://www.emro.who.int/countries.html). Only 'high coverage' sequences were selected, i.e. "entries with <1% Ns and <0.05% unique amino acid mutations (not seen in other sequences in the database) and no insertion/deletion unless verified by submitter". These were then analyzed by CoVsurver enabled by GISAID (https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/) and compared with reference strain hCoV-19/Wuhan/WIV04/2019 (accession

number: EPI_ISL_402124). Sequence metadata corresponding to age, gender, and date of collection were extracted from EpiCoV™ 'nextmeta' file. For clade analysis, we used the Nextclade tool from the Nextstrain project (https://clades.nextstrain.org/) [17].

Country Data and Mortality Rate

Data on population size, total COVID-19 cases, and total deaths were retrieved from the Worldometers website as of 14 November 2020 (https://www.worldometers.info/coronavirus/). Data on daily cases and deaths per country were extracted from WHO Coronavirus Disease (COVID-19) Dashboard (https://covid19.who.int/) [4]. CFR was calculated and adjusted for a 13-day lag time from reporting to death, such that the denominator was designated as the cumulative cases from 13 days earlier, while the numerator totaled the cumulative deaths till 13 days earlier plus half the additional deaths recorded during the lag period [18].

Statistical Analysis

The data were checked for completeness, and responses were coded and entered into the Statistical Package for the Social Sciences (SPSS) software version 20 for Windows, which was later used for statistical analyses [IBM: Statistical Package for the Social Sciences. SPSS Statistics 2013, 20.]. Eight of the 15 EMR countries that have sequence data available were excluded from cross-country fatality correlations for the following reasons: Egypt had a relatively high CFR (5.9%) probably as a result of the low number of reported cases (1073 cases per million) likely due to limited testing capacity; Iran had only 13 complete sequences (>25 kbp) out of 93 sequences and was thus less comparable to other sequences; Iraq had only 1 complete out of 31 sequences (**Supplemental Figure 1**); Pakistan, Tunisia, Lebanon, Kuwait, and Qatar had a relatively low sequence count each (<20). For the summary of the data, descriptive statistics were presented to

summarize the study variables of interest as counts and percentages across countries. Pie charts and bar plots were used to chart comparison in count and percentages of variants across all countries. An independent t-test was used to chart the comparison of variants via gender (male vs female), while one-way ANOVA was carried out for age distribution (0-9, 10-18, 19-49, 50-59, and >60). Pearson's correlation coefficients were used to examine the association between variant percentages and CFR for the seven countries considered in this study [19]. For all analyses done, a *p-value* of less than 0.05 was used to detect statistical significance.

Results

We identified 1414 high-coverage SARS-CoV-2 genome sequences from EMR countries – no sequences were available for Afghanistan, Djibouti, Libya, Somalia, Sudan, Syria, and Yemen. Four countries accounted for 73.7% of the sequence count overall – Saudi Arabia (n=521), UAE (n=186), Oman (n=185) and Egypt (n=150) (**Figure 1A**). Bahrain had the highest sequence to population ratio (44.1 in 1 million), while Oman had the highest number of sequences per reported COVID-19 cases (15.4 per 10,000 cases) (**Figure 1B**). When dates of collection were considered, sequences from the top three represented countries (i.e., Saudi Arabia, UAE, and Oman) were most evenly distributed, while those from Egypt showed a more abrupt pattern of sequence collection (**Figure 1C**).

Using CoVsurver from GISAID, amino acid substitutions were identified in each viral protein sequence and the top ten prevalent variants for each country were then plotted (excluding single-incidence variants) (**Figure 2**). In total, 1010 non-synonymous mutations were detected in EMR countries, out of which ten common variants were found among the majority of sequences: NSP12_P323L, S_D614G, NS3_Q57H, N_R203K, N_G204R, N_S194L, NSP6_L37F, NSP2_V198L, NS8_L84S and N_S202N. Besides, many country-specific substitutions were

detected (here defined as found in 3 other countries or less, and not listed in other countries' most common mutations). For instance, NSP5_G71S was found in 67% of Oman's sequences, NSP12_G823S in 24% of Egypt's sequences, and NSP12_M666I in 75% of Palestine's sequences (**Figure 2**). NSP12_P323L and S_D614G were predominantly detected in 10 out of 15 countries. S_D614G was also dominant in Bahrain. In Kuwait and Iran, NSP6_L37F and NSP2_V198L were the most frequently reported variants.

Next, we used the Nextclade tool from Nextstrain to assign clades (GISAID nomenclature) to the sequences (**Figure 3A)** [20]. Of all sequences from EMR countries, 65.6% belonged to the G (20A per Nextstrain clades), GR (20B), and GH (20C) clades, which emerged following the introduction of the S_D614G variant (**Figure 3B**, [21]). Moreover, these clades were almost exclusively represented as early as February and March in Saudi Arabia. In UAE and Egypt clade distribution was heterogeneous with multiple clades cocirculating between February through August (**Figure 3C**). In Oman, multiple clades were introduced into the country in February and March, but GH predominated as of April.

More specimens were sequenced from males in the EMR (2.05:1 male-to-female in top four represented countries, independent t-test $p < 0.004$): Saudi Arabia (4.44:1), UAE (2.25:1), Oman (1.42:1) and Egypt (1.49:1) (**Figure 4A**). No statistically significant difference in gender distribution of the top represented SARS-CoV-2 variants were observed in these countries (independent t-test for all variants, $p > 0.270$) (**Figure 6B**). By analyzing the age distribution of the top variants in these countries, we observed a statistically significant underrepresentation in the number of sequences of children (0-9 years) (3.43% ± 1.21 S.E.M.) and young adolescents (10-18 years) (2.26% ± 1.16), while adults (19-49) were mostly represented (62.5% ± 8.84) (ANOVA; $F = 73.4$, $p = 1.15E-9$) (**Figure 4C**). For the three remaining age categories (19-49, 50-

59 and >60 years), we report no statistically significant difference in the occurrence of the Wuhan-like or any of the identified variants (multivariate ANOVA for all variants, $F < 0.906$, $p > 0.487$) (**Figure 4D**)

We next sought to determine whether a correlation exists between the incidence of the ten most common variants and countries' CFRs. Only seven countries were included in this analysis – Saudi Arabia, Oman, UAE, Bahrain, Palestine, Morocco, and Jordan, where none of the variants was significantly correlated with an increase in CFR (**Figure 5**). Consistent with previous studies [22, 23], we confirm that NSP12_P323L seems to have quickly co-evolved following S_D614G introduction, as they are more often co-detected among SARS-CoV-2 genome sequences (**Figure 6A**). This double variant 'D614G & P323L' is found to be evenly distributed among different age groups in the top four represented countries (**Figure 6B**). After repeating the correlation analysis while accounting for the single versus double occurrence of D614G and P323L, we found that neither single nor double variants were significantly correlated with CFR (**Figure 6C**).

Finally, since sequences of Saudi Arabia and Oman were preferentially collected before July and those of Morocco and Jordan before May (**Figure 1C**), we reran our analysis but this time accounting for the cumulative incidence of the 'D614G & P323L' double variant and the CFR at the end of each month. We observed no association between these two variables from March through September (**Supplementary Figure 2**). To further investigate this correlation, we considered the cumulative monthly incidence of 'D614G & P323L' and the monthly CFRs (**Figure 7A**), where Saudi Arabia, Oman, and Egypt showed a positive correlation between the incidence of 'D614G & P323L' and the monthly CFR ($R=0.772$, $P=0.126$; $R=0.921$, $P=0.079$; $R=0.771$, $P=0.072$ respectively). In contrast, UAE had a significantly negative correlation between 'D614G & P323L' incidence and the monthly CFR ($R=-0.824$, $P=0.023$). This suggests that by accounting

for the change in CFR for each country with time the correlation between 'D614G & P323L' incidence and higher fatality does not necessarily hold true in contrary to previous findings [11, 19, 24].

Discussion

In this study, we analyzed whole-genome sequences of SARS-CoV-2 collected in the EMR countries. Significant data gaps are notable, whereby sequences were lacking for seven EMR countries – the majority of which are in war conflict zones (e.g., Syria, Libya, and Yemen [25]). In general, large disparities in the availability of sequences are observed across EMR. Only four countries reported >100 sequences spanning several months of the pandemic, with Saudi Arabia exceeding 500 sequence submissions. These data gaps highlight the challenges for monitoring the pandemic in developing countries particularly in conflict zones. It also emphasizes the importance of building regional sequencing capacities for timely and continuous monitoring of viral evolution during the pandemic [26, 27].

Despite the diversity in the substitutions detected across countries, only a handful of variants predominated. For instance, a mere 6 out of 260 non-synonymous mutations (2.3%) in Saudi Arabia and 5 out of 138 mutations (3.6%) in Oman were detected in more than 10% of the respective total of sequences (**Figure 2**). This suggests that consistent with other studies, most SARS-CoV-2 mutations are neutral or deleterious [10]. Although a random founder's effect cannot be ruled out, the increased frequency of certain substitutions might be attributable to a fitness advantage [12, 23, 28-30]. For instance, the S_D614G was found to predominate all viral strains globally (78% of sequences by May 2020) and was associated with increased viral loads in the upper respiratory tract of COVID-19 patients [31]. Along the same line, Plante et al. recently demonstrated that S_D614G enhanced viral replication and infectivity of SARS-CoV-2 in human

lung epithelial cells and primary human airway tissues and increased infectious virus loads in hamsters' nasal cavities [12]. Hou et al. further showed that the S_D614G variant displayed faster transmission and increased competitive fitness *in vivo* compared to wild-type virus in hamsters [32]. These findings were also confirmed in hACE2 knock-in mice as well as in ferrets [33]. On a molecular level, S_D614G allows for more efficient infectivity by shifting the spike protein to a more 'open' conformation that is more favorable of ACE2 binding [30].

Several studies reported a positive correlation between S_D614G incidence and mortality (CFR) [11, 19, 24]. However, a more recent study of 25,000 SARS-CoV-2 sequences in the United Kingdom did not associate the incidence of this mutation with higher COVID-19 mortality or with clinical severity in patients [34]. Similarly, Korber et al. did no find an association between S_D614G and increased disease severity or hospitalization [31]. In our sequence set, we saw no association between any of the ten most common mutations, including S_D614G, and CFRs of select EMR countries (**Figure 5**). However, our data as well as others' [22, 23] show that the S_D614G is frequently coupled with another substitution, NSP12_P323L, suggesting co-evolution of these variants (**Figure 6A**). In contrast to the well-studied S_D614G, the effect of NSP12_P323L on viral fitness is much less understood [35, 36]. One study speculated that it negatively impacts NSP8's association with RdRp leading to an attenuated polymerase activity [22], while another suggests it enhances hydrophobic NSP8-RdRp interactions leading to improved processivity of RdRp and enhanced viral replication overall [23]. In all cases, it is speculated that the 'epidemiological success' of the GH, G, and GR lineages could be attributed to the combined viral fitness advantage conferred by both NSP12_P323L and S_D614G [22]. Therefore, it makes more sense to consider the combined effect of the 'D614G & P323L' double variant on mortality and disease outcomes. Nonetheless, we found no correlation between the

cumulative incidence of the 'D614G & P323L' variant and CFRs in the EMR countries (**Figure 6C**), which confirms earlier observations that found no association between D614G & P323L and deaths per million, CFR and patient status [22].

Considering the limitations due to the discrepant number of sequences available (as mentioned above) or to country-specific variations in managing the COVID-19 pandemic (such as testing capacity, age stratification, enforcing/relaxing lockdown policies, etc.), we assessed the association of 'D614G & P323L' incidence with CFR within each of the top four represented countries over time. Interestingly, we found a positive, albeit non-significant, correlation between 'D614G & P323L' incidence and COVID-19 CFR in Saudi Arabia, Oman, and Egypt, but a statistically significant negative correlation in UAE (**Figure 7B**). This suggests that the reported correlation between these variants and increased CFR might be a mere coincidence rather than true causation. Noteworthy, our method of calculating daily/monthly CFR is credited for accounting for a 13-day lag between cumulative cases and deaths [18]. Yet, it overestimates the case fatality risk early on in the pandemic [37], a limitation that would apply to our final analysis. Then again, considering the overwhelming discrepancy in fatality correlations between countries (especially in UAE versus Oman with similar sequence counts (**Figure 7B**)), it is unlikely that early bias in CFR estimation would have altered our conclusion. Additionally, a more accurate method to estimate COVID-19 mortality would be the infection fatality rate (IFR), inferred from seroprevalence studies. However, such data is only available for Iran, Pakistan, and Qatar out of all EMR countries according to a recent WHO review [38].

On a final note, other genomic mutations were associated with milder disease, such as a major deletion (Δ382; nucleotide positions: 27,848 to 28,229), which results in truncated ORF7b and arrested transcription of ORF8 [39]. Patients infected with the SARS-CoV-2 Δ382 variant had less

systemic release of proinflammatory cytokines and better clinical outcomes (i.e. lower frequency of hypoxia requiring supplemental oxygen) than wild-type [40]. This variant, however, was not present in any of our sequence cohort (**data not shown**). Additionally, we did not detect an S_N439K variant, which enhances spike binding to its protein hACE2 and results in immune escape from a panel of neutralizing monoclonal antibodies [41].

Overall, our analysis reveals the diversity of SARS-CoV-2 in EMR and emphasizes the importance of considering temporal incidence and concurrent mutations in disease correlations. Our study is limited by the relatively low number of specimens sequenced per country spanning ten months of COVID-19 pandemic. There is a critical need for continuous and consistent surveillance of SARS-CoV-2 genomic variations in the EMR region.

Acknowledgments

S.O. is the recipient of a joint doctoral fellowship from the American University of Beirut (AUB) and the National Council for Scientific Research of Lebanon (CNRS-L). We gratefully acknowledge the authors from the originating laboratories responsible for obtaining the specimens, as well as the submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based (a detailed list of authors and laboratories can be found in the supplementary material).

Competing Interests: The authors declare no conflicts of interest.

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

References

[1] E. Callaway, Russia's fast-track coronavirus vaccine draws outrage over safety, Nature, 584 (2020) 334-335.
[2] H.V. Fineberg, The Toll of COVID-19, JAMA, 324 (2020) 1502-1503.
[3] J.P.A. Ioannidis, Global perspective of COVID-19 epidemiology for a full-cycle pandemic, Eur J Clin Invest, (2020) e13423.
[4] WHO, WHO Coronavirus Disease (COVID-19) Dashboard.
[5] WHO, Eastern Mediterranean Regional Office COVID-19 Dashboard.
[6] S. Su, G. Wong, W. Shi, J. Liu, A.C.K. Lai, J. Zhou, W. Liu, Y. Bi, G.F. Gao, Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses, Trends Microbiol, 24 (2016) 490-502.
[7] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science, 367 (2020) 1260-1263.
[8] J. Shang, Y. Wan, C. Luo, G. Ye, Q. Geng, A. Auerbach, F. Li, Cell entry mechanisms of SARS-CoV-2, Proc Natl Acad Sci U S A, 117 (2020) 11727-11734.
[9] E. Hartenian, D. Nandakumar, A. Lari, M. Ly, J.M. Tucker, B.A. Glaunsinger, The molecular virology of coronaviruses, J Biol Chem, 295 (2020) 12910-12934.
[10] L. van Dorp, M. Acman, D. Richard, L.P. Shaw, C.E. Ford, L. Ormond, C.J. Owen, J. Pang, C.C.S. Tan, F.A.T. Boshier, A.T. Ortiz, F. Balloux, Emergence of genomic diversity and recurrent mutations in SARS-CoV-2, Infect Genet Evol, 83 (2020) 104351.
[11] M. Eaaswarkhanth, A. Al Madhoun, F. Al-Mulla, Could the D614 G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality?, International Journal of Infectious Diseases, (2020).
[12] J.A. Plante, Y. Liu, J. Liu, H. Xia, B.A. Johnson, K.G. Lokugamage, X. Zhang, A.E. Muruato, J. Zou, C.R. Fontes-Garfias, D. Mirchandani, D. Scharton, J.P. Bilello, Z. Ku, Z. An, B. Kalveram, A.N. Freiberg, V.D. Menachery, X. Xie, K.S. Plante, S.C. Weaver, P.Y. Shi, Spike mutation D614G alters SARS-CoV-2 fitness, Nature, (2020).
[13] N. Vankadari, Overwhelming mutations or SNPs of SARS-CoV-2: A point of caution, Gene, 752 (2020) 144792.
[14] C. Wang, Z. Liu, Z. Chen, X. Huang, M. Xu, T. He, Z. Zhang, The establishment of reference sequence for SARS-CoV-2 and variation analysis, J Med Virol, 92 (2020) 667-674.
[15] M. Sallam, N. Ababneh, D. Dababseh, F. Bakri, A. Mahafzah, Temporal increase in D614G mutation of SARS-CoV-2 in the Middle East and North Africa: Phylogenetic and mutation analysis study, medRxiv, (2020) 2020.2008.2024.20176792.
[16] Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data - from vision to reality, Euro Surveill, 22 (2017).
[17] J. Hadfield, C. Megill, S.M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R.A. Neher, Nextstrain: real-time tracking of pathogen evolution, Bioinformatics, 34 (2018) 4121-4123.
[18] N. Wilson, A. Kvalsvig, L.T. Barnard, M.G. Baker, Case-Fatality Risk Estimates for COVID-19 Calculated by Using a Lag Time for Fatality, Emerg Infect Dis, 26 (2020) 1339-1441.
[19] Y. Toyoshima, K. Nemoto, S. Matsumoto, Y. Nakamura, K. Kiyotani, SARS-CoV-2 genomic variations associated with mortality rate of COVID-19, J Hum Genet, (2020).
[20] G.I.o.S.A.I.D. (GISAID), Clade and lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses, in, GISAID, Munich, 2020.

[21] E. Alm, E.K. Broberg, T. Connor, E.B. Hodcroft, A.B. Komissarov, S. Maurer-Stroh, A. Melidou, R.A. Neher, A. O'Toole, D. Pereyaslov, W.H.O.E.R.s. laboratories, G.E. group, W.H.O.E.R.s. laboratories, G.E. group*, Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020, Euro Surveill, 25 (2020).
[22] S. Ilmjärv, F. Abdul, S. Acosta-Gutiérrez, C. Estarellas, I. Galdadas, M. Casimir, M. Alessandrini, F.L. Gervasio, K.-H. Krause, Epidemiologically most successful SARS-CoV-2 variant: concurrent mutations in RNA-dependent RNA polymerase and spike protein, medRxiv, (2020).
[23] S.R. Kannan, A.N. Spratt, T.P. Quinn, X. Heng, C.L. Lorson, A. Sonnerborg, S.N. Byrareddy, K. Singh, Infectivity of SARS-CoV-2: there Is Something More than D614G?, J Neuroimmune Pharmacol, (2020).
[24] M. Becerra-Flores, T. Cardozo, SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate, Int J Clin Pract, 74 (2020) e13525.
[25] M.A. Daw, Corona virus infection in Syria, Libya and Yemen; an alarming devastating threat, Travel Med Infect Dis, 37 (2020) 101652.
[26] P. Salameh, COVID-19 in the Eastern Mediterranean Region: testing frequency, cumulative cases and mortality analysis, East Mediterr Health J, 26 (2020) 1005-1010.
[27] A.M.O. Al Saidi, F.A. Nur, A.S. Al-Mandhari, M. El Rabbat, A. Hafeez, A. Abubakar, Decisive leadership is a necessity in the COVID-19 response, The Lancet, 396 (2020) 295-298.
[28] E. Callaway, The coronavirus is mutating - does it matter?, Nature, 585 (2020) 174-177.
[29] N.D. Grubaugh, W.P. Hanage, A.L. Rasmussen, Making Sense of Mutation: What D614G Means for the COVID-19 Pandemic Remains Unclear, Cell, 182 (2020) 794-795.
[30] L. Yurkovetskiy, X. Wang, K.E. Pascal, C. Tomkins-Tinch, T.P. Nyalile, Y. Wang, A. Baum, W.E. Diehl, A. Dauphin, C. Carbone, K. Veinotte, S.B. Egri, S.F. Schaffner, J.E. Lemieux, J.B. Munro, A. Rafique, A. Barve, P.C. Sabeti, C.A. Kyratsous, N.V. Dudkina, K. Shen, J. Luban, Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant, Cell, 183 (2020) 739-751 e738.
[31] B. Korber, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya, B. Foley, K.M. Hastie, M.D. Parker, D.G. Partridge, C.M. Evans, T.M. Freeman, T.I. de Silva, C.-G.G. Sheffield, C. McDanal, L.G. Perez, H. Tang, A. Moon-Walker, S.P. Whelan, C.C. LaBranche, E.O. Saphire, D.C. Montefiori, Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus, Cell, 182 (2020) 812-827 e819.
[32] Y.J. Hou, S. Chiba, P. Halfmann, C. Ehre, M. Kuroda, K.H. Dinnon, 3rd, S.R. Leist, A. Schafer, N. Nakajima, K. Takahashi, R.E. Lee, T.M. Mascenik, R. Graham, C.E. Edwards, L.V. Tse, K. Okuda, A.J. Markmann, L. Bartelt, A. de Silva, D.M. Margolis, R.C. Boucher, S.H. Randell, T. Suzuki, L.E. Gralinski, Y. Kawaoka, R.S. Baric, SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo, Science, (2020).
[33] B. Zhou, T.T.N. Thao, D. Hoffmann, A. Taddeo, N. Ebert, F. Labroussaa, A. Pohlmann, J. King, J. Portmann, N.J. Halwe, L. Ulrich, B.S. Trueb, J.N. Kelly, X. Fan, B. Hoffmann, S. Steiner, L. Wang, L. Thomann, X. Lin, H. Stalder, B. Pozzi, S. de Brot, N. Jiang, D. Cui, J. Hossain, M. Wilson, M. Keller, T.J. Stark, J.R. Barnes, R. Dijkman, J. Jores, C. Benarafa, D.E. Wentworth, V. Thiel, M. Beer, SARS-CoV-2 spike D614G variant confers enhanced replication and transmissibility, bioRxiv, (2020).

[34] E.M. Volz, V. Hill, J.T. McCrone, A. Price, D. Jorgensen, A. O'Toole, J.A. Southgate, R. Johnson, B. Jackson, F.F. Nascimento, Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity, MedRxiv, (2020).
[35] H.S. Hillen, G. Kokic, L. Farnung, C. Dienemann, D. Tegunov, P. Cramer, Structure of replicating SARS-CoV-2 polymerase, Nature, 584 (2020) 154-156.
[36] M. Pachetti, B. Marini, F. Benedetti, F. Giudici, E. Mauro, P. Storici, C. Masciovecchio, S. Angeletti, M. Ciccozzi, R.C. Gallo, D. Zella, R. Ippodrino, Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant, J Transl Med, 18 (2020) 179.
[37] Y. Ge, S. Sun, Estimation of Coronavirus Disease Case-Fatality Risk in Real Time, Emerg Infect Dis, 26 (2020) 1922-1923.
[38] J.P.A. Ioannidis, Infection fatality rate of COVID-19 inferred from seroprevalence data, Bulletin of the World Health Organization, (2020).
[39] Y.C.F. Su, D.E. Anderson, B.E. Young, M. Linster, F. Zhu, J. Jayakumar, Y. Zhuang, S. Kalimuddin, J.G.H. Low, C.W. Tan, W.N. Chia, T.M. Mak, S. Octavia, J.M. Chavatte, R.T.C. Lee, S. Pada, S.Y. Tan, L. Sun, G.Z. Yan, S. Maurer-Stroh, I.H. Mendenhall, Y.S. Leo, D.C. Lye, L.F. Wang, G.J.D. Smith, Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2, mBio, 11 (2020).
[40] B.E. Young, S.W. Fong, Y.H. Chan, T.M. Mak, L.W. Ang, D.E. Anderson, C.Y. Lee, S.N. Amrun, B. Lee, Y.S. Goh, Y.C.F. Su, W.E. Wei, S. Kalimuddin, L.Y.A. Chai, S. Pada, S.Y. Tan, L. Sun, P. Parthasarathy, Y.Y.C. Chen, T. Barkham, R.T.P. Lin, S. Maurer-Stroh, Y.S. Leo, L.F. Wang, L. Renia, V.J. Lee, G.J.D. Smith, D.C. Lye, L.F.P. Ng, Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study, Lancet, 396 (2020) 603-611.
[41] E.C. Thomson, L.E. Rosen, J.G. Shepherd, R. Spreafico, A. da Silva Filipe, J.A. Wojcechowskyj, C. Davis, L. Piccoli, D.J. Pascall, J. Dillen, S. Lytras, N. Czudnochowski, R. Shah, M. Meury, N. Jesudason, A. De Marco, K. Li, J. Bassi, A. O'Toole, D. Pinto, R.M. Colquhoun, K. Culap, B. Jackson, F. Zatta, A. Rambaut, S. Jaconi, V.B. Sreenu, J. Nix, R.F. Jarrett, M. Beltramello, K. Nomikou, M. Pizzuto, L. Tong, E. Cameroni, N. Johnson, A. Wickenhagen, A. Ceschi, D. Mair, P. Ferrari, K. Smollett, F. Sallusto, S. Carmichael, C. Garzoni, J. Nichols, M. Galli, J. Hughes, A. Riva, A. Ho, M.G. Semple, P.J. Openshaw, K. Baillie, S.J. Rihn, S.J. Lycett, H.W. Virgin, A. Telenti, D. Corti, D.L. Robertson, G. Snell, The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity, bioRxiv, (2020) 2020.2011.2004.355842.

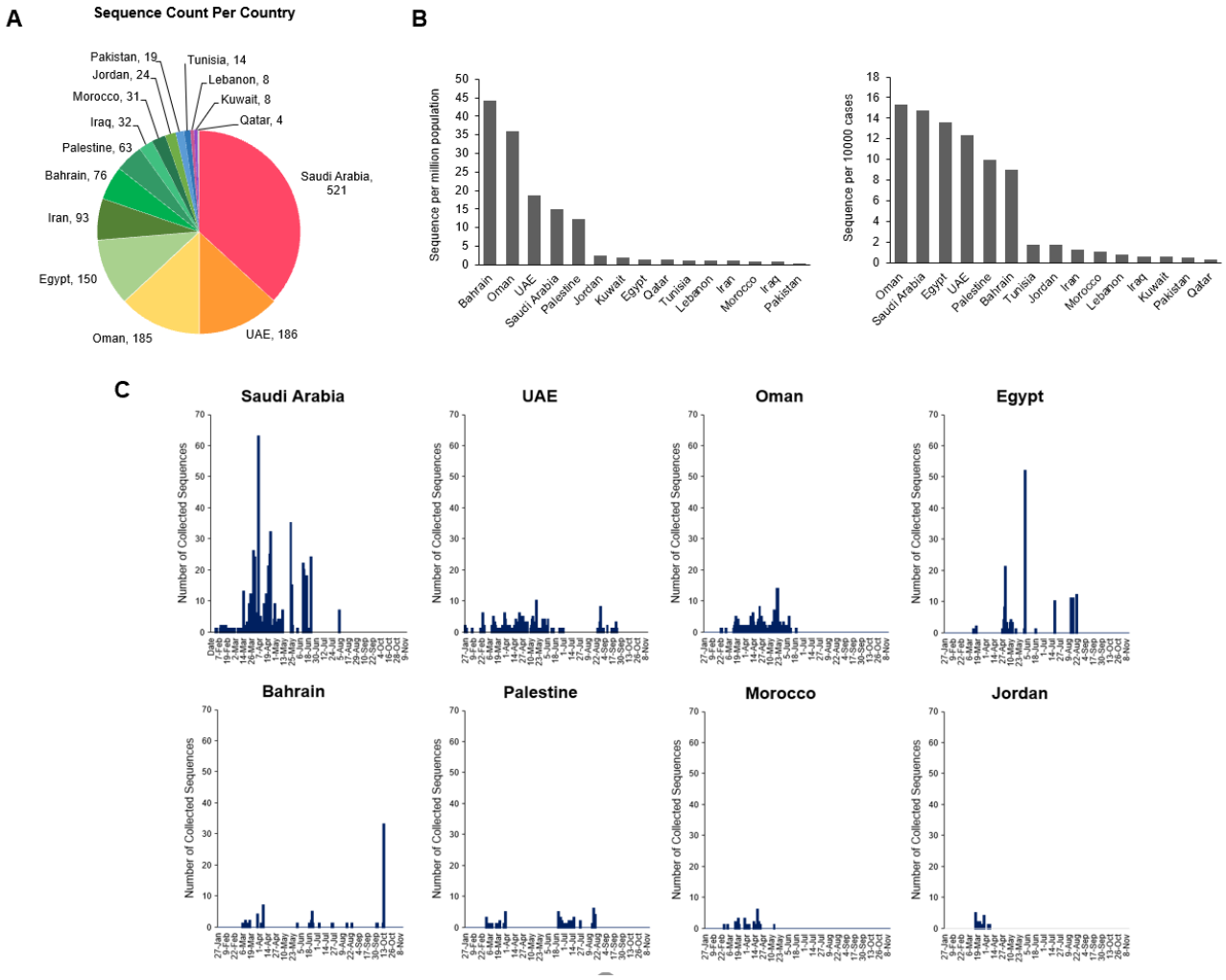


Figure 1. Distribution of SARS-CoV-2 sequences in EMR countries. (A) Pie chart showing number of available sequences as of 12 November 2020. (B) Histograms showing, left, ratios of sequences per million population and, right, sequences per 10,000 confirmed cases of SARS-CoV-2. (C) Temporal distribution of sequence collection in eight EMR countries.

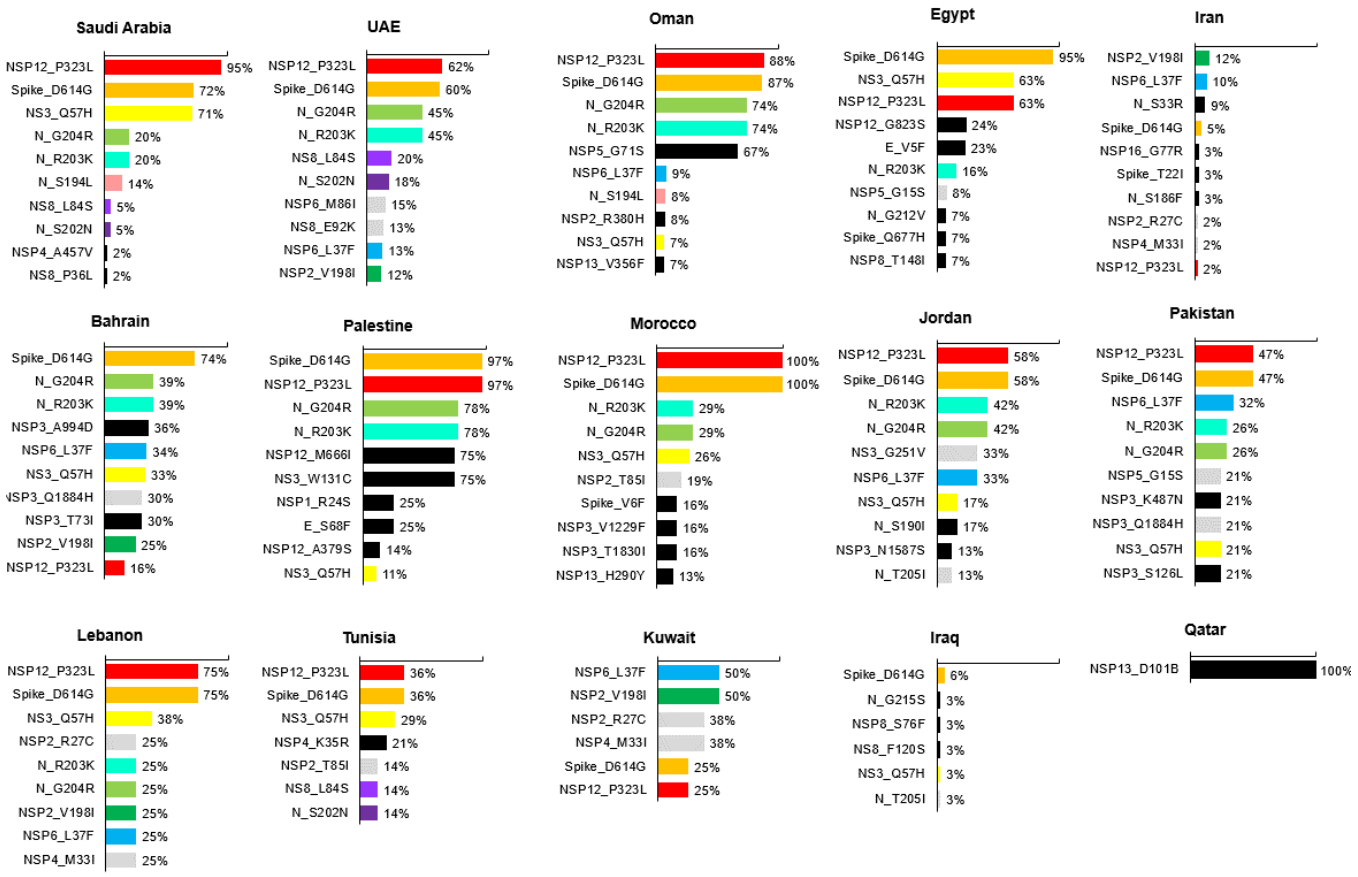


Figure 2. Common SARS-CoV-2 substitutions detected in EMR countries. Bar graphs showing percentages of most prevalent SARS-CoV-2 variants per country, excluding single-incidence variants. Colored are the top ten most represented substitutions, while in black are country-specific ones.

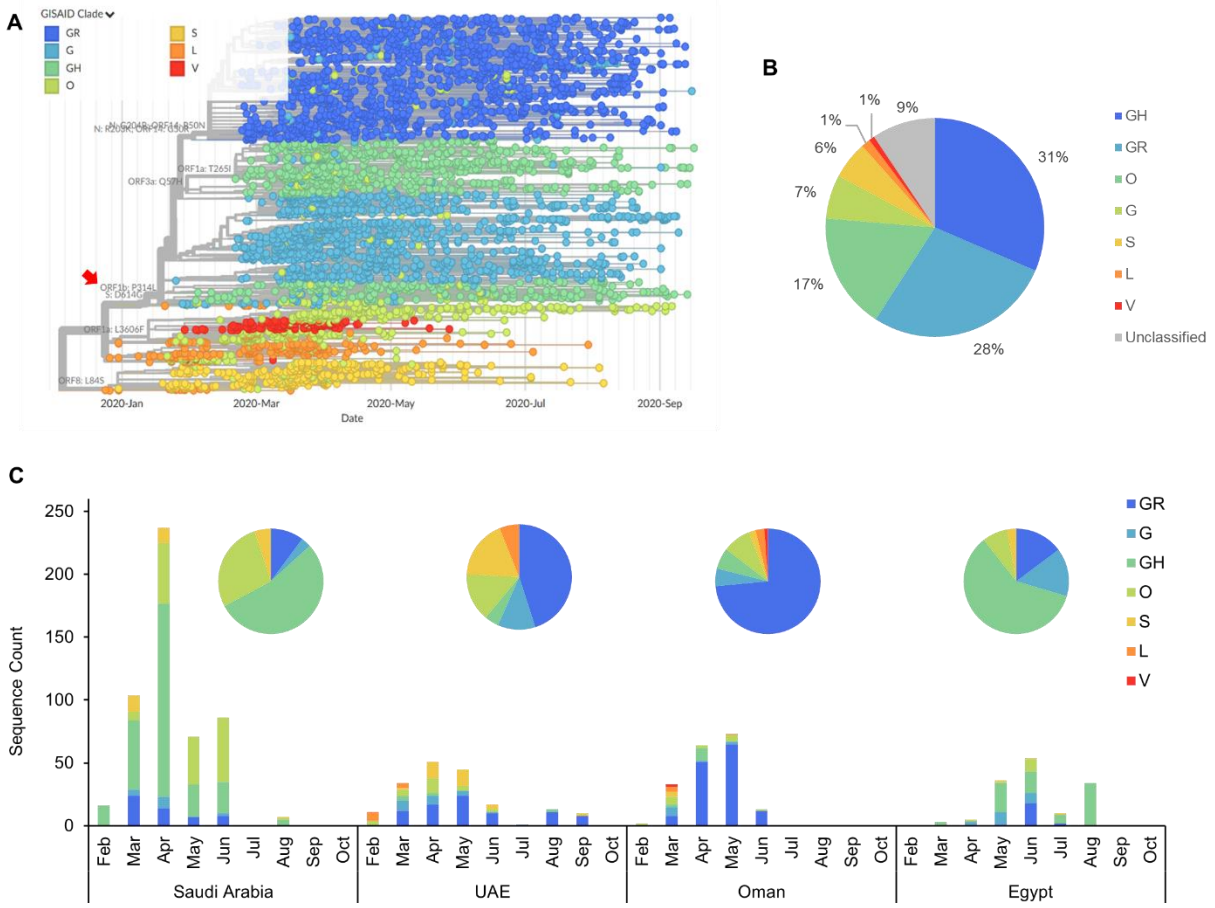


Figure 3. EMR sequences are dominantly distributed in G, GR and GH GISAID clades. (A) Phylogenetic tree of GISAID clades plotted over time. Red arrow points to the introduction of S_D614G, marking the emergence of G, GR and GH clades (reproduced from Nextstrain [Ref. 17]. (B) Clade distribution of all 1414 sequences. (C) Temporal distribution of sequences as clades in the top four represented countries.

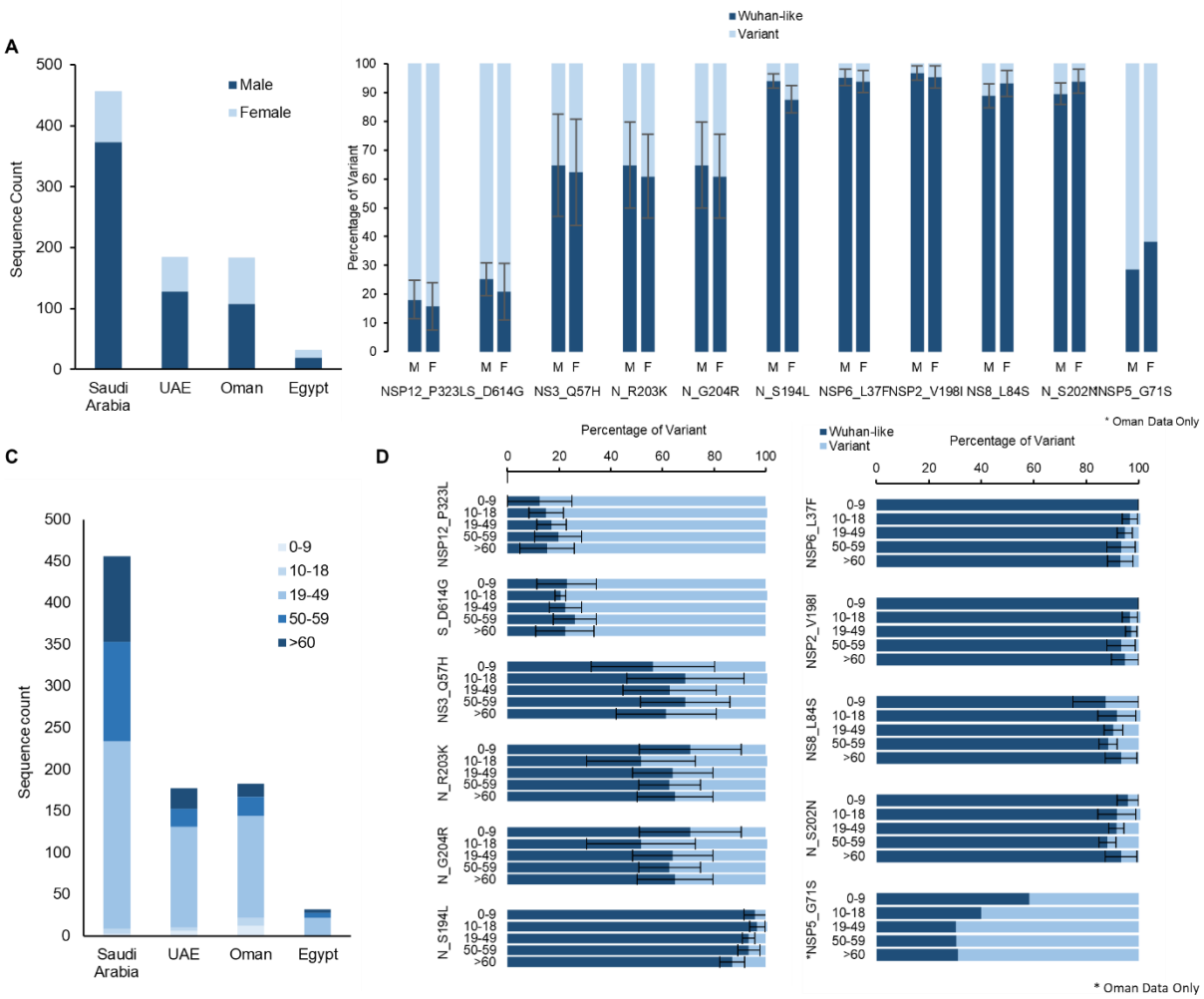


Figure 4. Gender and age distribution of SARS-CoV-2 variants in top four represented EMR countries. Gender distribution of (A) sequences of Saudi Arabia, UAE, Oman and Egypt and (B) select top most frequent variants of these countries. (C) Age groups distribution of sequences from top four EMR countries and (D) of most frequent variants. NSP5_G71S was specific to Oman.

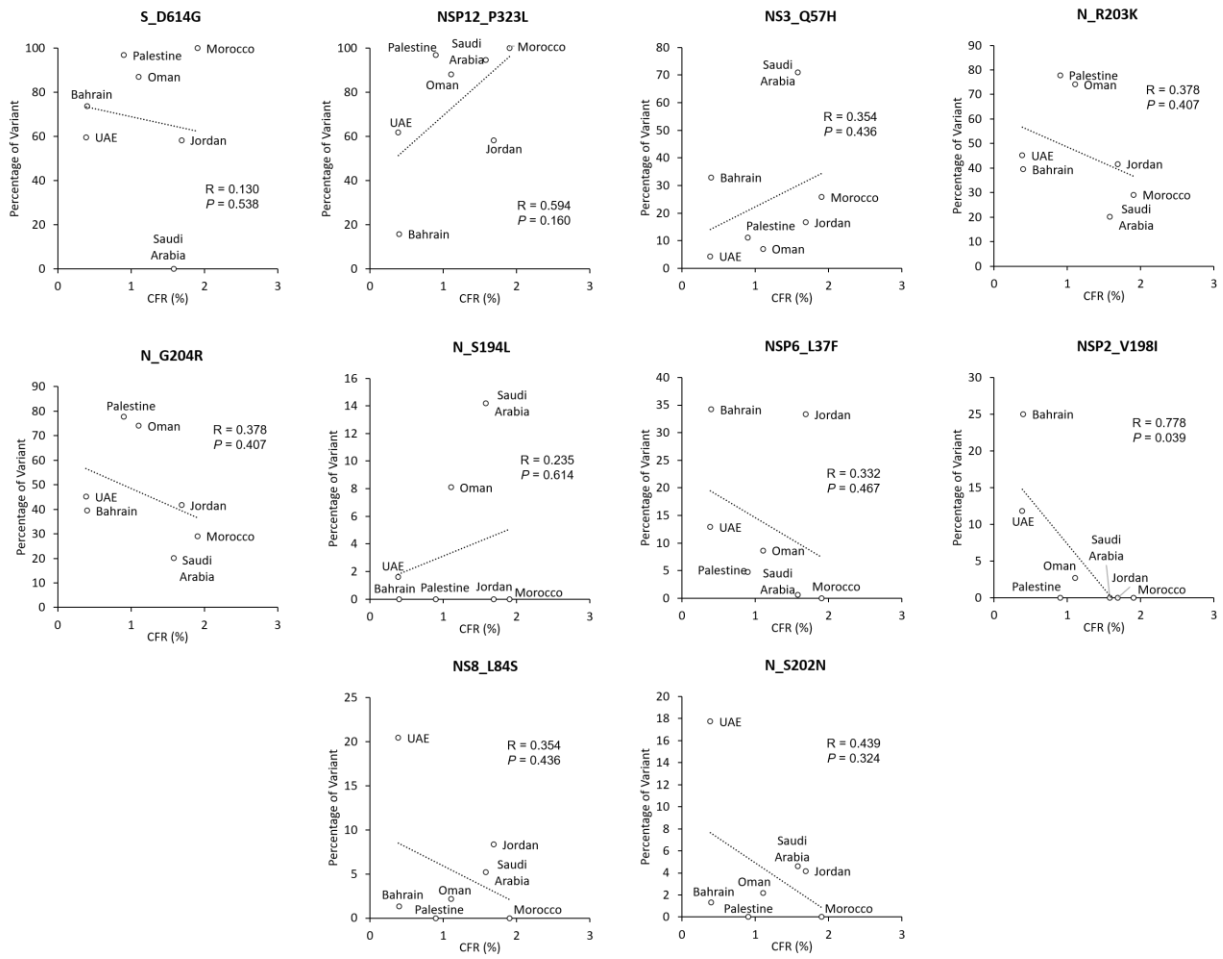


Figure 5. No correlation of single SARS-CoV-2 variants with case fatality rates in seven EMR countries. Scatter plots showing percentages of each of the top ten represented substitutions with cumulative CFR as of 12 November 2020. R, Pearson's correlation coefficient.

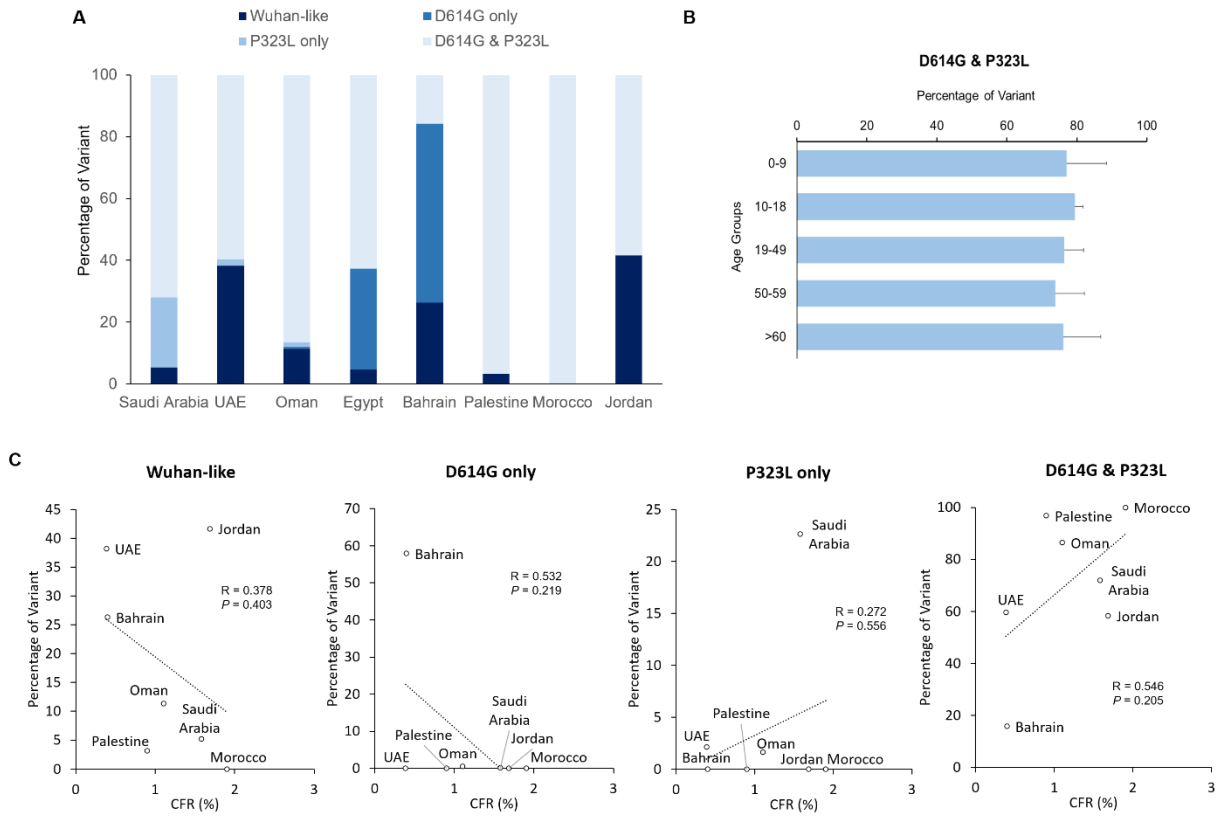


Figure 6. The 'D614G & P323L' double variant is predominant in select EMR countries, but does not correlate with increased fatality. (A) Percentages of Wuhan-like, S_D614G only, NSP_P323L only and 'D614G & P323L' double variants in eight EMR countries. (B) Distribution of the 'D614 & P323L' variant based on age groups in top four represented EMR countries. (C) Scatter plots showing correlation of four variants with CFR in seven EMR countries. R, Pearson's correlation coefficient.

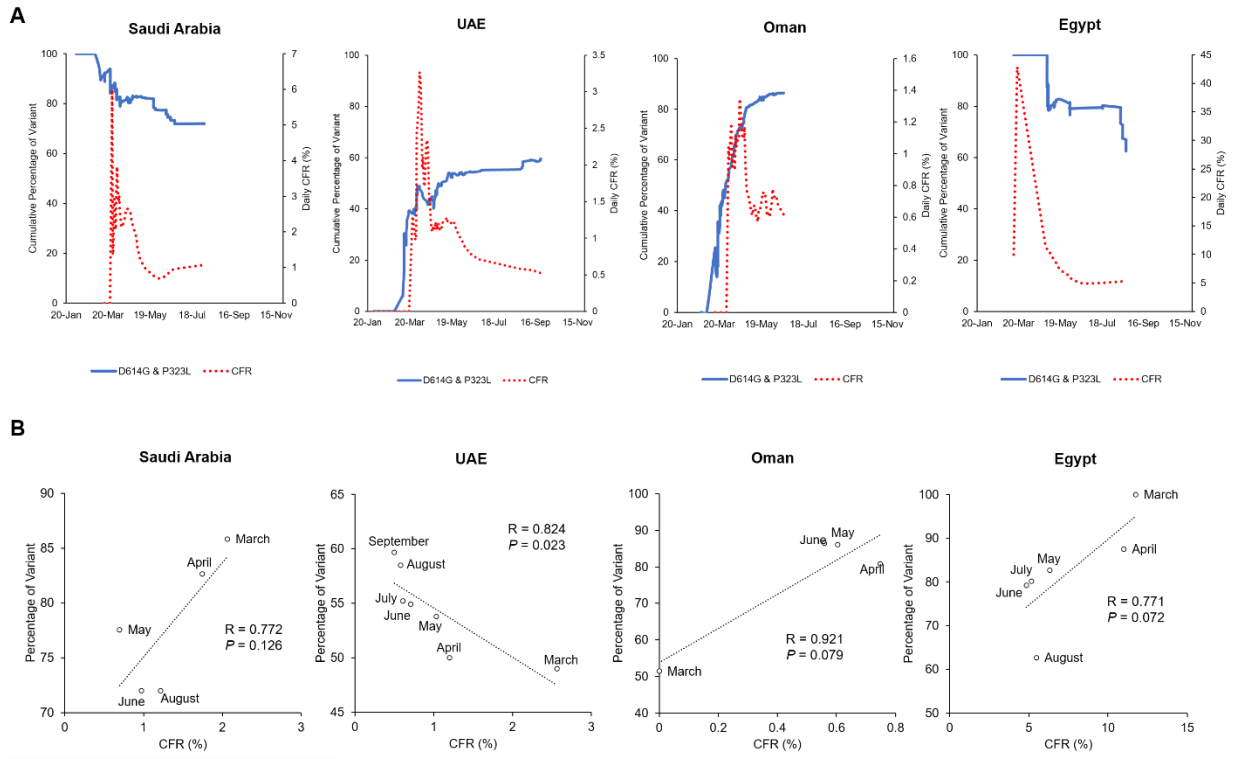
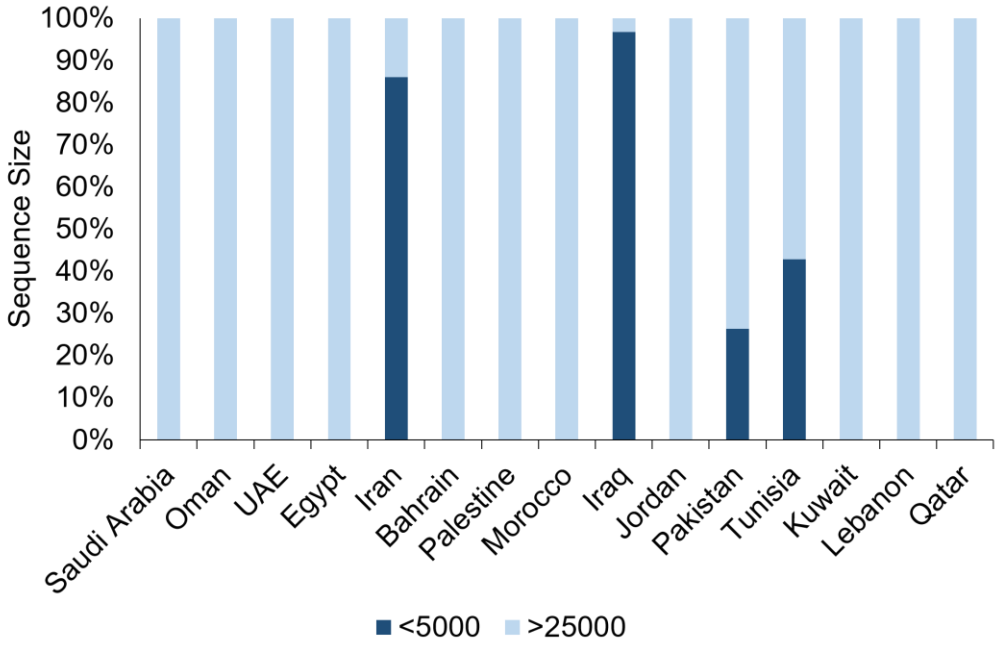
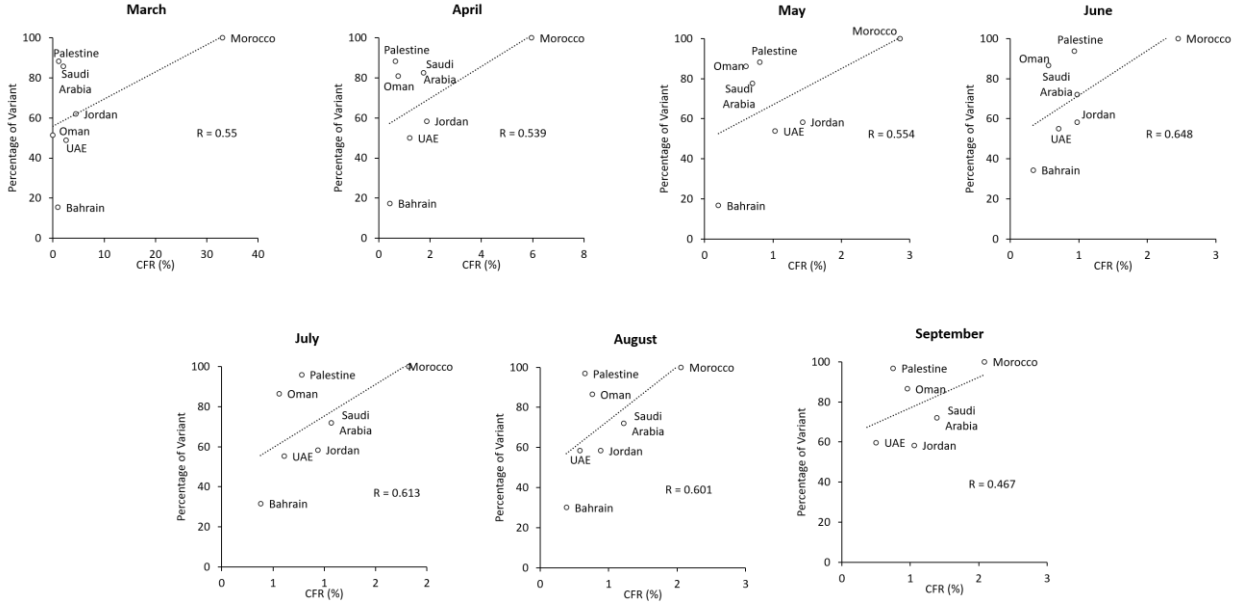


Figure 7. Temporal distribution and fatality correlation of the 'D614G & P323L' double variant with CFR in each of the top four represented EMR Countries. (A) Time-based evolution of double variant percentages and CFR in Saudi Arabia, UAE, Oman and Egypt. (B) Scatter plots showing correlations of the 'D614G & P323L' double variant with CFR over select months of pandemic life in the top four EMR countries. R, Pearson's correlation coefficient.



Supplementary Figure 1. Genome length of available EMR sequences.



Supplementary Figure 2. Correlation between the 'D614G & P323L' double variant with CFR over seven months of the pandemic. Scatter plots showing cumulative percentages of the double variant in seven EMR countries versus respective CFR as of the end of each month, March through September. R, Pearson's correlation coefficient.