



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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Research paper

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

Saad Omais^a, Samer Kharroubi^b, Hassan Zaraket^{c,d,*}

^a Department of Biology, Faculty of Arts and Sciences, American University of Beirut, Beirut, Lebanon

^b Department of Nutrition and Food Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut, Beirut, Lebanon

^c Faculty of Medicine, Center for Infectious Diseases Research, American University of Beirut, Beirut 1107 2020, Lebanon

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



ARTICLE INFO

Keywords:

SARS-CoV-2

Whole genome

Variants

Case fatality rate

East Mediterranean Region

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 reported in the Eastern Mediterranean Region (EMR) countries, as of 1 January 2021. The majority (~75%) of these sequences originated from three out of 22 EMR countries, and 65.8% of all sequences belonged to GISAID clades GR, GH, G and GV. A delay ranging between 30 and 150 days from sample collection to sequence submission was observed across all countries, limiting the utility of such data in informing public health policies. We identified ten common non-synonymous mutations represented among SARS-CoV-2 in the EMR and several country-specific ones. Two substitutions, spike_D614G and NSP12_P323L, were predominantly concurrent in most countries. While the single incidence of NSP12_P323L was positively correlated with higher case fatality rates in EMR, no such association was established for the double (spike_D614G and NSP12_P323L) concurrent variant across the region. Our study identified critical data gaps in EMR highlighting the importance of enhancing surveillance and sequencing capacities in the region.

1. 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 (Keaten, 2020). Since its declaration as a pandemic seven months earlier, COVID-19 has inflicted substantial health and economic burden around the world (Fineberg, 2020; Ioannidis, 2020). As of 22 January 2021, there have been over 96 million confirmed cases and 2 million deaths attributed to COVID-19 (WHO, 2021a). Specifically, 5,461,398 confirmed cases (of whom 4,815,552 have recovered) and 130,079 deaths were reported in the Eastern Mediterranean Region (EMR) (WHO, 2021b).

Prior to SARS-CoV-2 emergence, six human coronavirus (HCoV) strains were known to infect humans: two of which caused severe

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) (Su et al., 2016). 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 (Wrapp et al., 2020). 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 (Shang et al., 2020). 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). ORF_a and -b constitute nearly two-thirds of the genome and

Abbreviations: WHO, World Health Organization; EMR, Eastern Mediterranean Region; HCoV, human coronavirus; S, spike; hACE2, human angiotensin-converting enzyme 2; RBD, receptor-binding domain; ORF, open reading frames; NSP, non-structural proteins; RdRp, RNA-dependent RNA polymerase; N, nucleocapsid; M, membrane; E, envelope; CFR, case fatality rate; SPSS, Statistical Package for the Social Sciences; UAE, United Arab Emirates; AUB, American University of Beirut; CNRS-L, National Council for Scientific Research of Lebanon.

* Corresponding author at: Department of Experimental Pathology, Faculty of Medicine, Immunology & Microbiology, American University of Beirut, Beirut 1107 2020, Lebanon.

E-mail address: hz34@aub.edu.lb (H. Zaraket).

<https://doi.org/10.1016/j.gene.2021.145843>

Received 12 January 2021; Received in revised form 25 March 2021; Accepted 13 July 2021

Available online 16 July 2021

0378-1119/© 2021 Elsevier B.V. All rights reserved.

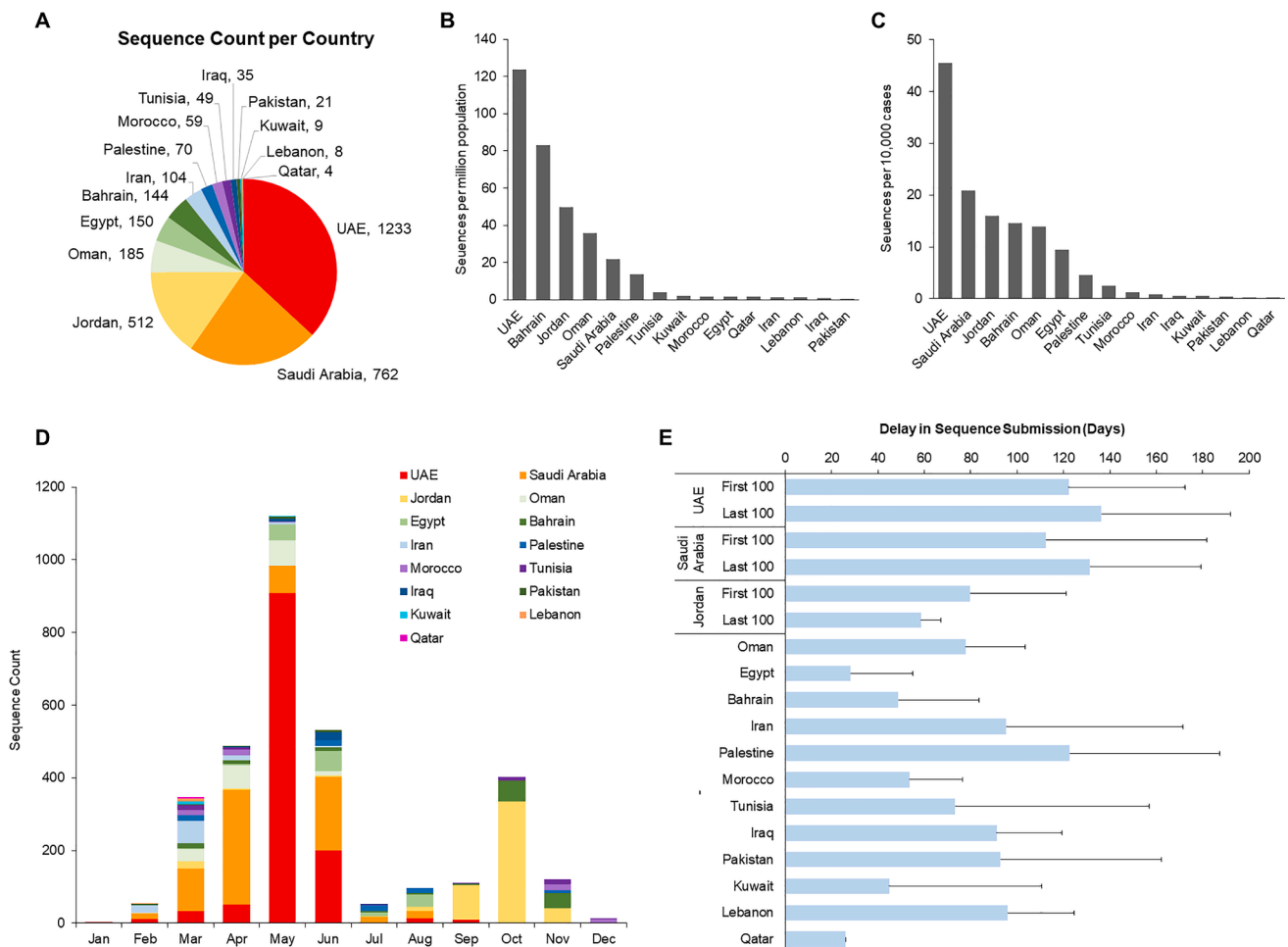


Fig. 1. Distribution of SARS-CoV-2 sequences in EMR countries. (A) Pie chart showing number of available sequences as of 1 January 2021. (B) Histogram showing ratio of sequences per million population. (C) Histogram showing ratio of sequences per 10,000 confirmed cases of SARS-CoV-2. (D) Temporal distribution of sequence collection in EMR countries from January through December 2020. (E) Averages of delay in sequence submission (in days) of available EMR countries. For UAE, Saudi Arabia and Jordan, only the first and last 100 collected sequences were considered. Error bars represent standard deviation.

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 (Hartenian et al., 2020).

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 6×10^{-4} nucleotides/genome/year (van Dorp et al., 2020a). 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 (van Dorp et al., 2020a; Easwar-khanth et al., 2020; Plante et al., 2021; Vankadari, 2020; Wang et al., 2020). Specifically, Sallam et al. reported an increase in the incidence of S_{D614G} variant between February and June 2020 in the Middle East and North Africa (Sallam et al., 2021). 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.

2. Methods

2.1. Sequence retrieval and analysis

SARS-CoV-2 sequences were downloaded from GISAID EpiCoVTM database (Shu and McCauley, 2017), which were submitted by 1 January 2021 from 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 and gender were extracted from EpiCoVTM 'nextmeta' file for sequences submitted as of 12 November 2020. For clade analysis, we used the Nextclade tool from the Nextstrain project (<https://clades.nextstrain.org/>) (Hadfield et al., 2018) and GISAID clade designation provided by CoVsurver analysis.

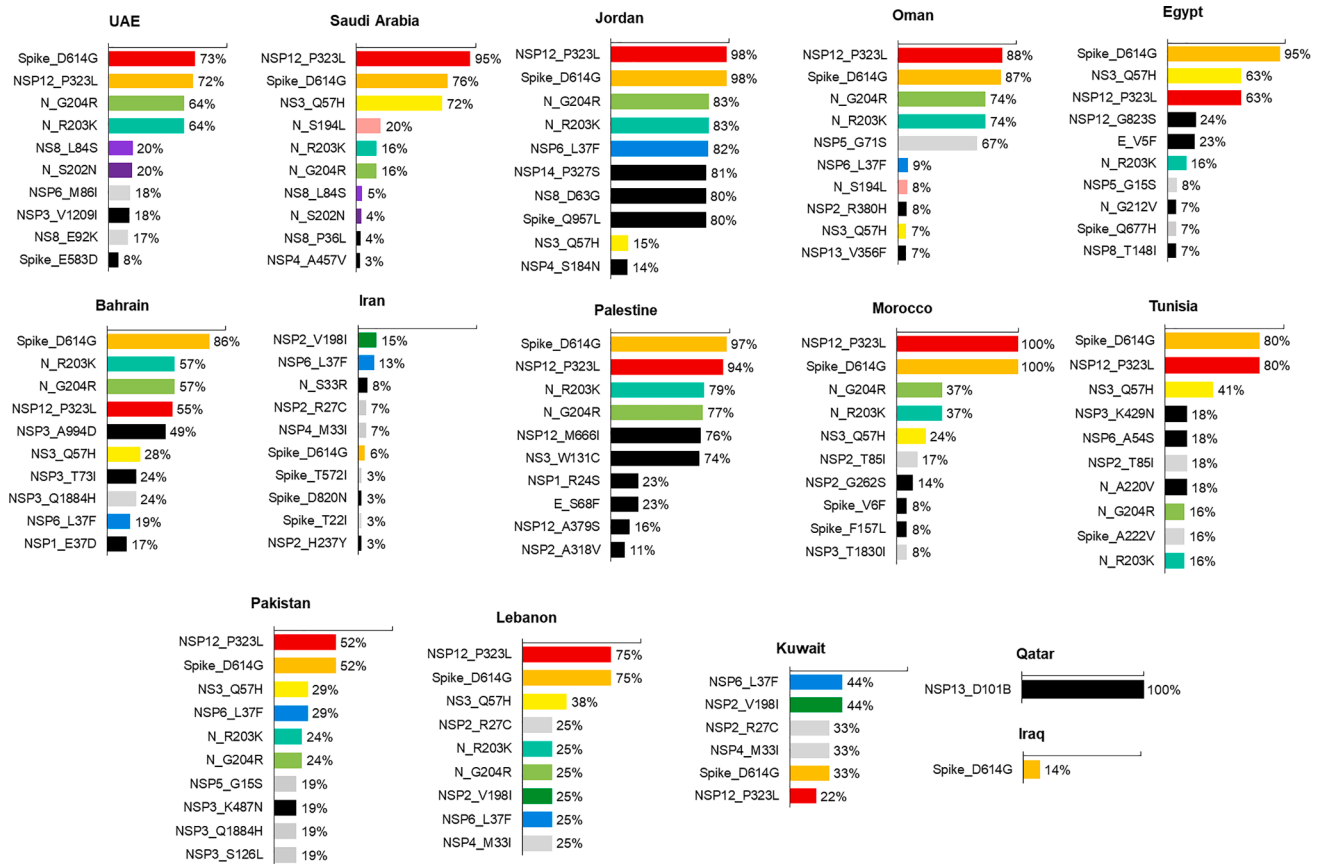


Fig. 2. Common SARS-CoV-2 substitutions detected in EMR countries. Bar graphs showing percentages of most prevalent SARS-CoV-2 variants per country, excluding single-incidence variants. Colored are the top ten most represented substitutions; in black are country-specific ones.

3. Country data and mortality rate

Data on population size, total COVID-19 cases, and total deaths were retrieved from the Worldometers website as of 22 January 2021 (<http://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/>) (WHO, 2021a). 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 (Wilson et al., 2020).

3.1. Statistical analysis

The data were checked for completeness 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; Iraq, Pakistan, Tunisia, Lebanon, Kuwait, and Qatar had a relatively low sequence count each (<50, 1.5% of all sequences) (Fig. 1A), Iran had only 18 complete sequences (>25 kbp) out of 104 sequences and was thus less comparable to other sequences (Supplementary Fig. 1). 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 (Toyoshima et al., 2020). For all analyses done, a *p*-value of <0.05 was used to detect statistical significance.

4. Results

We identified 3345 high-coverage SARS-CoV-2 genome sequences from EMR countries – no sequences were available for Afghanistan, Djibouti, Libya, Somalia, Sudan, Syria, and Yemen. Three countries accounted for ~75% of the sequence count overall – UAE (n = 1233), Saudi Arabia (n = 762) and Jordan (n = 512) (Fig. 1A). All but five EMR countries – Iran, Tunisia, Iraq, Pakistan and Kuwait – had near-full length sequences of SARS-CoV-2 (Supplementary Fig. 1). UAE had the highest sequence to population ratio (123.8 in 1 million), but also the highest number of sequences per reported COVID-19 cases (45.5 per 10,000 cases) (Fig. 1B and C). When dates of collection were considered, none of the countries showed routine and consistent sequencing; most sequences were collected before July (e.g. from UAE, Saudi Arabia, Oman) while only few countries (mainly Jordan, Bahrain and Morocco) reported sequences from the last three months (Fig. 1D). Moreover, a significant delay between sample collection and sequence submission was observed across all countries (ranging 30–150 days), with no noticeable progress between first and last 100 days for the top three represented countries (Fig. 1E).

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)

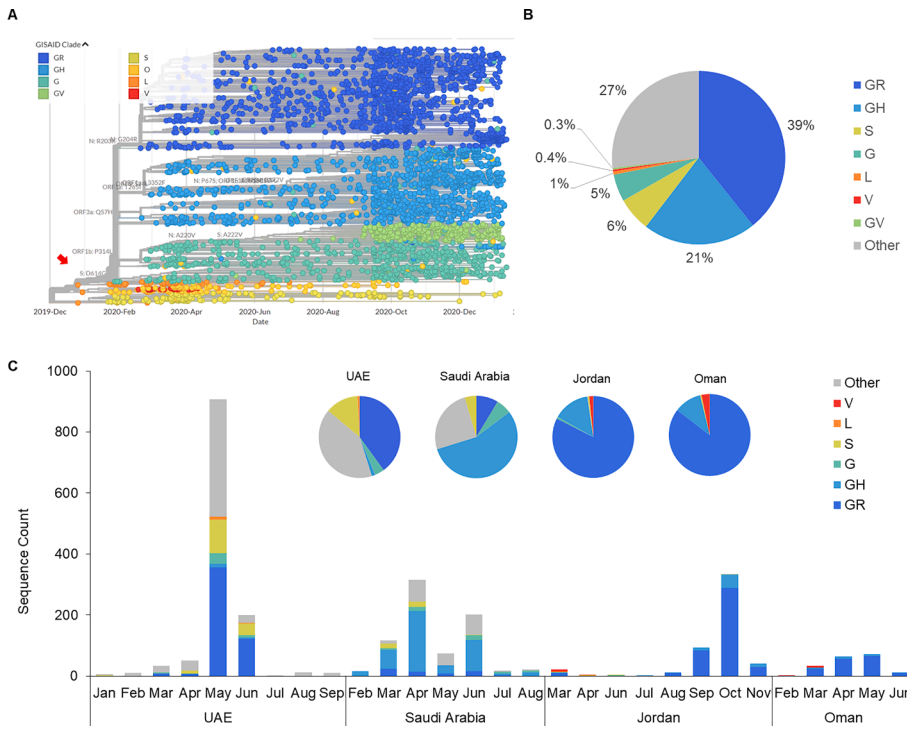


Fig. 3. EMR sequences are dominantly distributed in G, GR and GH GISAID clades. (A) Phylogenetic tree of GISAID clades plotted over time. Red arrow points to the introduction of S_D614G, marking the emergence of G, GR and GH clades (reproduced from Nextstrain). (B) Clade distribution of all 3345 sequences. (C) Temporal distribution of sequences as clades in the top four represented countries. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

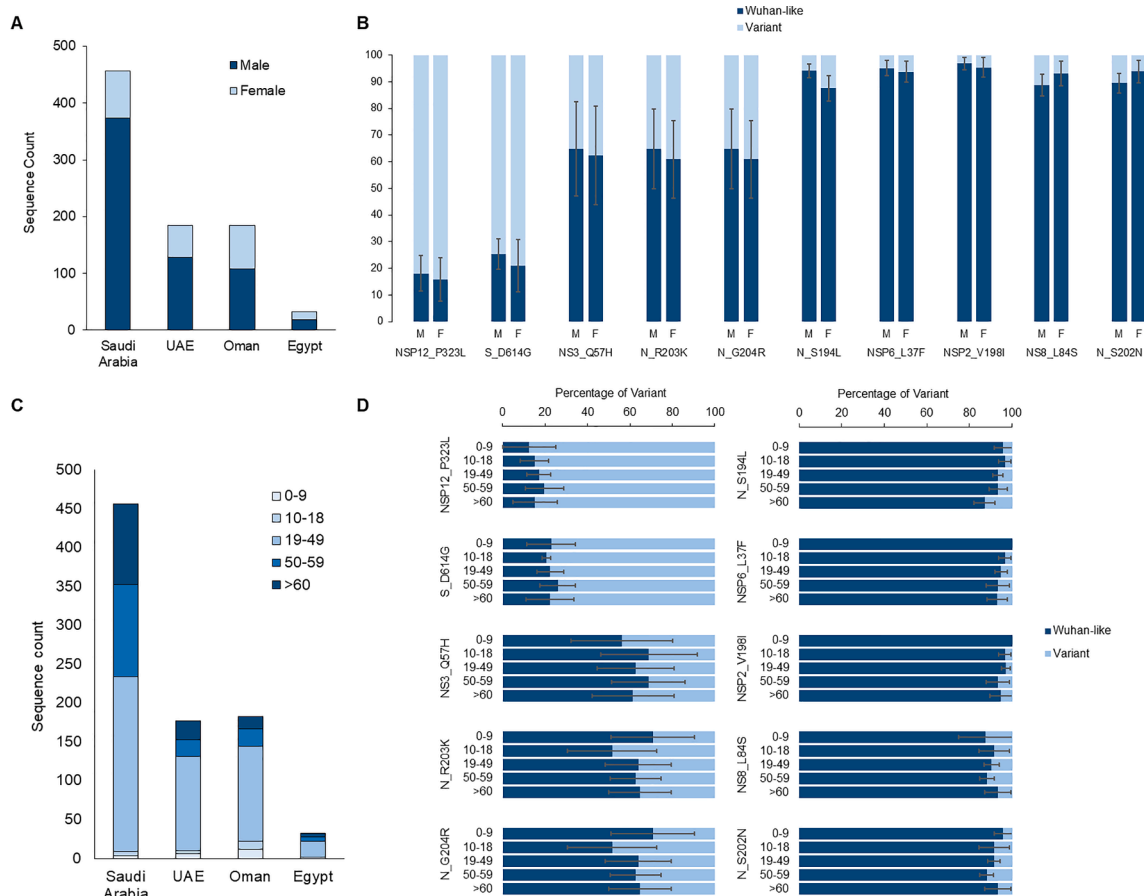


Fig. 4. Gender and age distribution of SARS-CoV-2 variants in four EMR countries. Gender distribution of (A) sequences of Saudi Arabia, UAE, Oman and Egypt and (B) select top most frequent variants of these countries. (C) Age groups distribution of sequences from four EMR countries and (D) of most frequent variants. Sequences in this figure are as of 12 November 2020 due to the inability to extract demographics from metafiles after this date.

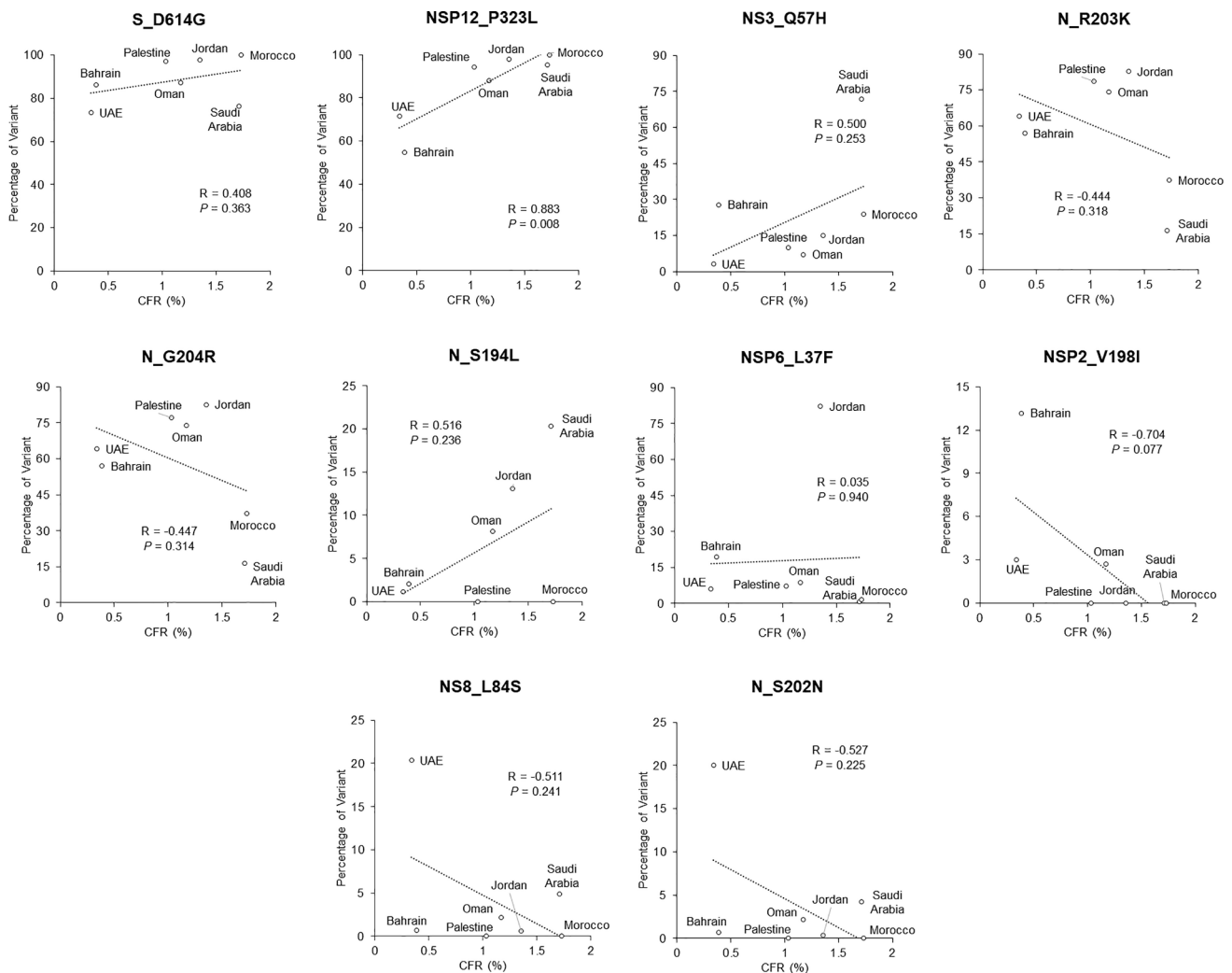


Fig. 5. Only NSP12_P323L of the ten most common SARS-CoV-2 single variants correlates with higher case fatality rates in seven EMR countries. Scatter plots showing percentages of each of the top ten represented substitutions with cumulative CFR as of 1 January 2021. R, Pearson's correlation coefficient.

(Fig. 2). In total, 2155 non-synonymous mutations were detected in EMR countries, out of which ten common variants were found to be relatively most common (i.e. in ≥ 8 EMR countries and $>2.5\%$ of all sequences). These are: 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, NSP3_V1209I was found in 18% of UAE's sequences, Spike_Q957L in 80% of Jordan's sequences and NSP12_M666I in 76% of Palestine's sequences (Fig. 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 (Fig. 3A) (GISAID, 2020). Of all sequences from EMR countries, 65.8% belonged to the G (20A per Nextstrain clades), GR (20B), GH (20C) and GV (20A.EU1) clades, which emerged following the introduction of the S_D614G variant (Fig. 3B, (Alm et al., 2020)). Moreover, these clades were almost exclusively represented as early as February and March in Saudi Arabia. In UAE, Jordan and Oman, clade distribution was heterogeneous with multiple clades cocirculating between February through November (Fig. 3C).

For the sequence cohort submitted by 12 November 2020, 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) (Fig. 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, $p > 0.270$) (Fig. 4B). By analyzing the age distribution of the top variants in these countries, we observed a statistically significant underrepresentation in the number of sequenced viruses from children (0–9 years) ($3.43\% \pm 1.21$ S.E. M.) and young adolescents (10–18 years) ($2.26\% \pm 1.16$), while adults (19–49) were more represented ($62.5\% \pm 8.84$) (ANOVA; $F = 73.4$, $p = 1.15E-9$) (Fig. 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, $F < 0.906$, $p > 0.487$) (Fig. 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 – UAE, Saudi Arabia, Jordan, Oman, Bahrain, Palestine and Morocco – where only NSP12_P323L out of the ten variants was significantly correlated with an increase in CFR ($R = 0.883$, $P = 0.008$) (Fig. 5). However, consistent with previous studies (Ilmjarv et al., 2021; Kannan et al., 2020), 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 (Fig. 6A). This double variant 'D614G & P323L' is found to be evenly represented among different age groups in UAE, Saudi Arabia, Oman and Egypt (for sequences submitted by 12

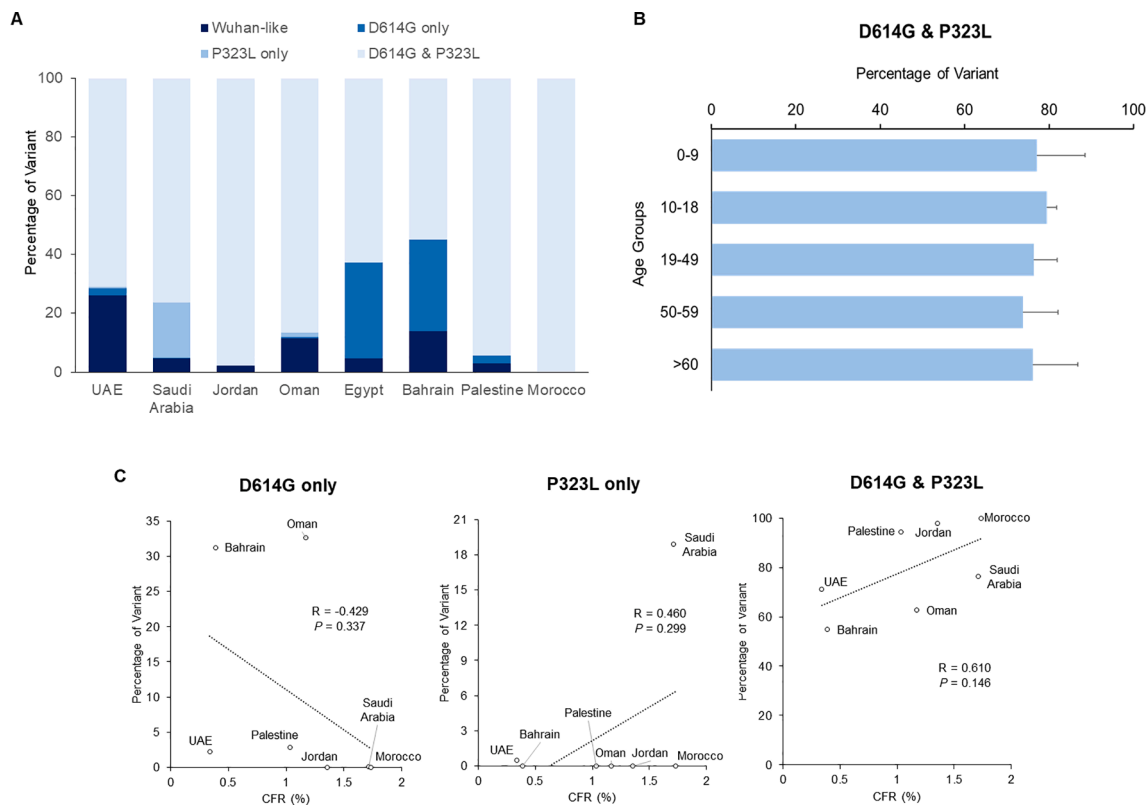


Fig. 6. The 'D614G & P323L' double variant is predominant in select EMR countries, but does not correlate with increased fatality. (A) Percentages of Wuhan-like, S_{D614G} only, NSP12_P323L only and 'D614G & P323L' double variants in eight EMR countries. (B) Distribution of the 'D614G & P323L' variant based on age groups in top four represented EMR countries. (C) Scatter plots showing correlation of three variants with CFR in seven EMR countries. R, Pearson's correlation coefficient.

November) (Fig. 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. The 'P323L only' variant was mainly detected in Saudi Arabia (19% of its sequences) and among 0.5% sequences reported from UAE, while in the other countries it was exclusively co-detected with D614G. As such, we suggest that the earlier correlation of NSP12 P323L is likely the result of considering single- and double-variants as one group, specifically in Saudi Arabia. However, when such distinction was made, no correlation with higher CFR was reported in either group (Fig. 6C).

Since sequences of UAE, Saudi Arabia and Oman were dominantly collected before September and those of Jordan and Bahrain were collected afterwards (Fig. 1D), we repeated 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 significant association between these two variables from March through September (Supplementary Fig. 2). To further investigate this correlation, we considered the cumulative monthly incidence of 'D614G & P323L' and the monthly CFRs (Fig. 7A): while UAE showed a significantly negative correlation between 'D614G & P323L' incidence and the monthly CFR ($R = -0.830$, $P = 0.021$), Saudi Arabia showed a significantly positive correlation ($R = 0.909$, $P = 0.012$). Meanwhile, Jordan and Oman had no significant associations ($R = -0.101$, $P = 0.849$; $R = 0.921$, $P = 0.079$ respectively) (Fig. 7B). 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 contrary to previous findings (Eaaswarkhanth et al., 2020; Toyoshima et al., 2020; Becerra-Flores and Cardozo, 2020).

5. 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 (Daw, 2020)). In general, large disparities in the availability of sequences are observed across EMR. Only five countries reported ≥ 150 sequences spanning the first year of the pandemic, with UAE, Saudi Arabia and Jordan 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 (Salameh, 2020; Al Saiti et al., 2020).

Despite the diversity in the substitutions detected across countries, only a handful of variants predominated. For instance, a mere 9 out of 945 non-synonymous mutations (0.95%) in UAE and 6 out of 421 mutations (1.4%) in Saudi Arabia were detected in $>10\%$ of the respective total of sequences (Fig. 2). This suggests that consistent with other studies, most SARS-CoV-2 mutations are neutral or deleterious (van Dorp et al., 2020a). Although a random founder's effect cannot be ruled out, the increased frequency of certain substitutions might be attributable to a fitness advantage (Grubaugh et al., 2020; Kannan et al., 2020; Plante et al., 2021; Zhang et al., 2020; Yurkovetskiy et al., 2020). 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 (Korber et al., 2020). 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

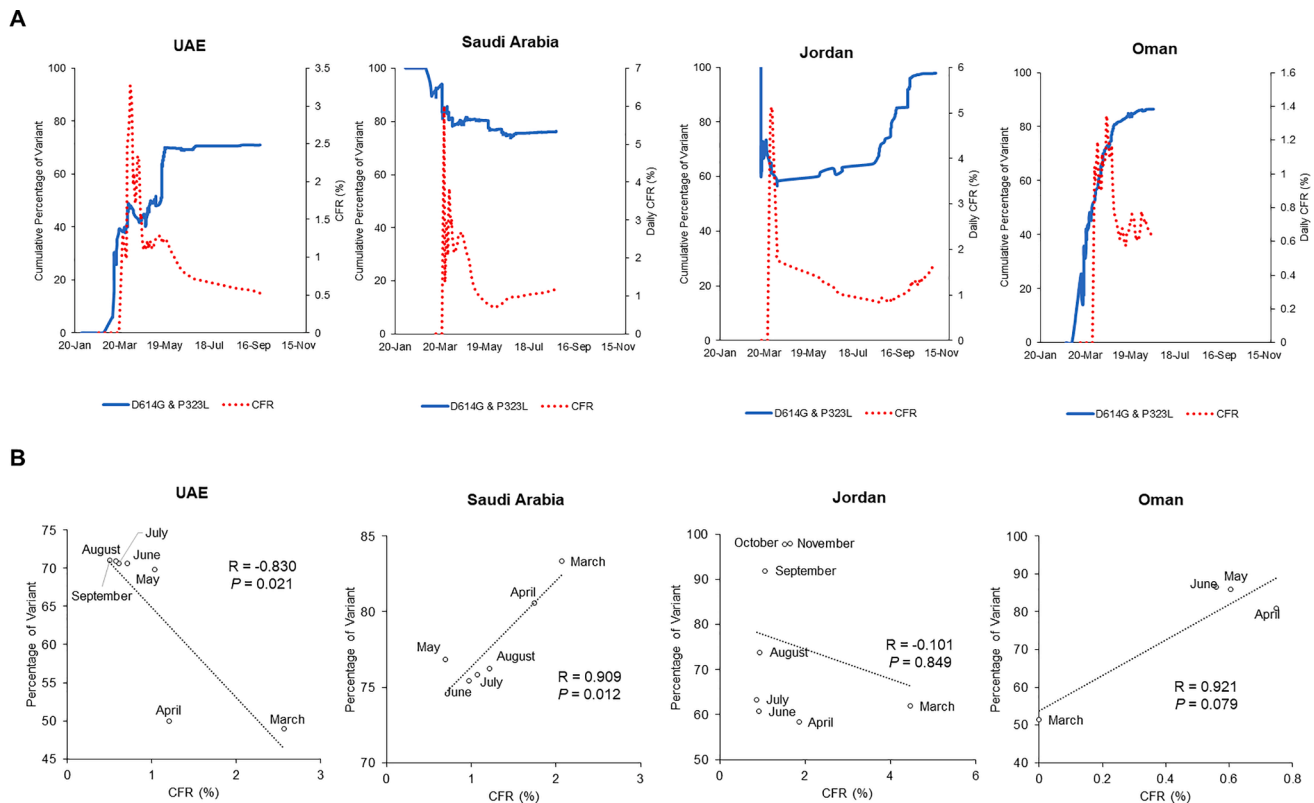


Fig. 7. Temporal distribution and fatality correlation of the 'D614G & P323L' double variant with CFR in each of the top four represented EMR Countries. (A) Time-based evolution of double variant percentages and CFR in Saudi Arabia, UAE, Oman and Egypt. (B) Scatter plots showing correlations of the 'D614G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

increased infectious virus loads in hamsters' nasal cavities (Plante et al., 2021). 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 (Hou et al., 2020). These findings were also confirmed in hACE2 knock-in mice as well as in ferrets (Zhou et al., 2021). 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 (Yurkovetskiy et al., 2020). Interestingly, however, a recent analysis by van Dorp et al. could not identify any single recurrent mutation, including the S_D614G, to be convincingly associated with increased viral transmission in humans (van Dorp et al., 2020b).

Several studies reported a positive correlation between S_D614G incidence and mortality (CFR) (Eaaswarkhanth et al., 2020; Toyoshima et al., 2020; Becerra-Flores and Cardozo, 2020). 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 (Volz et al., 2021). Similarly, Korber et al. did not find an association between S_D614G and increased disease severity or hospitalization (Korber et al., 2020). In our sequence set, we saw no association between S_D614G with CFRs of select EMR countries. Instead, the increasing incidence of NSP12_P323L (of the ten most common substitutions), when considered alone, was correlated with a higher mortality rate. However, our data as well as others' (Ilmjarv et al., 2021; Kannan et al., 2020) show that the NSP12_P323L and S_D614G are frequently coupled in most sequences, suggesting co-evolution of these variants (Fig. 6A). Therefore, we assessed the combined effect of the 'D614G & P323L' double variant on mortality. We found no correlation between the cumulative incidence of the 'D614G & P323L' variant and CFRs in the EMR countries, consistent with earlier observations that found no association between 'D614G & P323L' and deaths per million, CFR and patient status (Ilmjarv et al.,

2021). In contrast to the well-studied S_D614G, the effect of NSP12_P323L on viral fitness is much less understood (Hillen et al., 2020; Pachetti et al., 2020). One study speculated that it negatively impacts NSP8's association with RdRp leading to an attenuated polymerase activity (Ilmjarv et al., 2021), while another suggested that it enhances hydrophobic NSP8-RdRp interactions leading to improved processivity of RdRp and enhanced viral replication overall (Kannan et al., 2020). In all cases, it is speculated that the 'epidemiological success' of the G, GH, GR and GV lineages could be attributed to the combined viral fitness advantage conferred by both NSP12_P323L and S_D614G (Ilmjarv et al., 2021).

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 significantly negative correlation between 'D614G & P323L' incidence and COVID-19 CFR in UAE, but an otherwise positive correlation in Saudi Arabia (Fig. 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 (Wilson et al., 2020). Yet, it overestimates the case fatality risk early on in the pandemic (Ge and Sun, 2020), a limitation that would apply to our final analysis. Then again, considering the overwhelming discrepancy in fatality correlations between countries, 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 (Ioannidis, 2020).

Other genomic mutations were associated with milder disease, such as a major deletion (Δ 382; nucleotide positions: 27,848 to 28,229), which results in truncated ORF7b and arrested transcription of ORF8 (Su et al., 2020). Patients infected with the SARS-CoV-2 Δ 382 variant had less systemic release of proinflammatory cytokines and better clinical outcomes (i.e. lower frequency of hypoxia requiring supplemental oxygen) than wild-type (Young et al., 2020). This variant, however, was not present in any of our sequence cohort. Moreover, an S_N439K variant was reported to enhance spike binding to its protein hACE2, resulting in immune escape from a panel of neutralizing monoclonal antibodies (Thomson et al., 2021). We detected only one sequence carrying this mutation from Morocco (EPI_ISL_728353).

On a final note, the COVID-19 pandemic cannot be viewed in the absence of other respiratory viruses, especially in view of the increased concerns regarding a ‘Twindemic’ (Adlhoch et al., 2021; Uyeki et al., 2020). In the case of influenza, few cases were reported in EMR countries in 2020–21 season (<20 positive specimens per week; GISRS report (WHO Global Influenza Surveillance and Response System (GISRS), 2021), a pattern that was established worldwide (Jones, 2020). Previous studies from the United States and Australia reported a co-infection rate ranging from 2% to 6.9% between SARS-CoV-2 and other respiratory viruses including rhinovirus/enterovirus, RSV, other coronaviruses, and influenza virus (Kim et al., 2020; Kim et al., 2021). Elghoudi et al. reported an incidence of 3.1% (9/288) respiratory co-infections with SARS-CoV-2 in children and young adolescents in a hospital in UAE (Elghoudi et al., 2020). In addition to co-infections with respiratory viruses a study from Pakistan reported a relatively high incidence of co-infections between SARS-CoV-2 and the endemic dengue virus – 25% (5/20) of studied samples (Saddique et al., 2020). A meta-analysis study estimated that while 3% of COVID-19 patients (out of 1014 patients from 16 studies) had a confirmed viral co-infection, there was no significant increase in hospitalization or ICU admission in the co-infected cohort (Lansbury et al., 2020). Overall, the incidence of respiratory viral co-infections is low and may not be considered a key factor associated with mortality.

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 the COVID-19 pandemic. There is a critical need for continuous and consistent surveillance of SARS-CoV-2 genomic variations in the EMR region.

CRedit authorship contribution statement

Saad Omais: Methodology, Data curation, Visualization, Writing - original draft. **Samer Kharroubi:** Methodology, Writing - review & editing. **Hassan Zaraket:** Conceptualization, Supervision, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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). We also thank Dr. Colin A. Smith and Dr. Heinrich zu Dohnna (Biology department, American University of Beirut) for their valuable comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gene.2021.145843>.

References

- Fineberg, H.V., 2020. The Toll of COVID-19. *JAMA* 324, 1502–1503.
- Ilmjarv, S., Abdul, F., Acosta-Gutierrez, S., Estarellas, C., Galdadas, I., Casimir, M., Alessandrini, M., Gervasio, F.L., Krause, K.H., 2021. Concurrent mutations in RNA-dependent RNA polymerase and spike protein emerged as the epidemiologically most successful SARS-CoV-2 variant. *Sci Rep* 11, 13705.
- Ioannidis, J.P.A., 2020. Global perspective of COVID-19 epidemiology for a full-cycle pandemic. *Eur J Clin Invest* 50 (12). <https://doi.org/10.1111/eci.v50.1210.1111/eci.13423>.
- WHO, WHO Coronavirus Disease (COVID-19) Dashboard (2021). <https://covid19.who.int/>. Last Accessed: Jan 22 2021.
- WHO, Eastern Mediterranean Regional Office COVID-19 Dashboard (2021). <http://www.emro.who.int/index.html>. Last Accessed: Jan 22 2021.
- 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.
- Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., McLellan, J.S., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263.
- Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., Li, F., 2020. Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* 117, 11727–11734.
- Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J.M., Glaunsinger, B.A., 2020. The molecular virology of coronaviruses. *J Biol Chem* 295, 12910–12934.
- van Dorp, L., Acman, M., Richard, D., Shaw, L.P., Ford, C.E., Ormond, L., Owen, C.J., Pang, J., Tan, C.C.S., Boshier, F.A.T., Ortiz, A.T., Balloux, F., 2020a. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol* 83, 104351.
- Eaaswarkhanth, M., Al Madhoun, A., Al-Mulla, F., 2020. 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* 96, 459–460.
- Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., Mirchandani, D., Scharton, D., Billello, J.P., Ku, Z., An, Z., Kalveram, B., Freiberg, A.N., Menachery, V.D., Xie, X., Plante, K.S., Weaver, S.C., Shi, P.-Y., 2021. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592 (7852), 116–121.
- Vankadari, N., 2020. Overwhelming mutations or SNPs of SARS-CoV-2: A point of caution. *Gene* 752, 144792.
- Wang, C., Liu, Z., Chen, Z., Huang, X., Xu, M., He, T., Zhang, Z., 2020. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol* 92, 667–674.
- Sallam, M., Ababneh, N.A., Dababseh, D., Bakri, F.G., Mahafzah, A., 2021. Temporal increase in D614G mutation of SARS-CoV-2 in the Middle East and North Africa. *Heliyon* 7 (1), e06035. <https://doi.org/10.1016/j.heliyon.2021.e06035>.
- Shu, Y., McCauley, J., 2017. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Euro Surveill* 22.
- Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34, 4121–4123.
- Wilson, N., Kvalsvig, A., Barnard, L.T., Baker, M.G., 2020. Case-Fatality Risk Estimates for COVID-19 Calculated by Using a Lag Time for Fatality. *Emerg Infect Dis* 26, 1339–1441.
- Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., Kiyotani, K., 2020. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J Hum Genet* 65 (12), 1075–1082.
- 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).
- Kannan, S.R., Spratt, A.N., Quinn, T.P., Heng, X., Lorson, C.L., Sonnerborg, A., Byrreddy, S.N., Singh, K., 2020. Infectivity of SARS-CoV-2: there Is Something More than D614G? *J Neuroimmune Pharmacol*.
- Becerra-Flores, M., Cardozo, T., 2020. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int J Clin Pract* 74 (8). <https://doi.org/10.1111/ijcp.v74.810.1111/ijcp.13525>.
- Daw, M.A., 2020. Corona virus infection in Syria, Libya and Yemen; an alarming devastating threat. *Travel Med Infect Dis* 37, 101652. <https://doi.org/10.1016/j.tmaid.2020.101652>.
- Salameh, P., 2020. COVID-19 in the Eastern Mediterranean Region: testing frequency, cumulative cases and mortality analysis. *East Mediterr Health J* 26, 1005–1010.
- Al Saidi, A.M.O., Nur, F.A., Al-Mandhari, A.S., El Rabbat, M., Hafeez, A., Abubakar, A., 2020. Decisive leadership is a necessity in the COVID-19 response. *The Lancet* 396, 295–298.

- GISAID, 2020. Clade and lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses. GISAID, Munich.
- 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* 182, 794–795.
- L. Yurkovetskiy, X. Wang, K.E. Pascal, C. Tomkins-Tinch, T.P. Nyalile, Y. Wang, A. Baum, W.E. Diehl, A. Dauphin, C. Carbone, C. Veinotte, S.B. Egri, S.F. Schaffner, J.E. Lemieux, J.B. Munro, A. Raffique, 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.
- Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Peng, H., Quinlan, B.D., Rangarajan, E.S., Pan, A., Vanderheiden, A., Suthar, M.S., Li, W., Izard, T., Rader, C., Farzan, M., Choe, H., 2020. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun* 11, 6013.
- 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.
- Hou, Y.J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon 3rd, K.H., Leist, S.R., Schafer, A., Nakajima, N., Takahashi, K., Lee, R.E., Mascenik, T.M., Graham, R., Edwards, C.E., Tse, L.V., Okuda, K., Markmann, A.J., Bartelt, L., de Silva, A., Margolis, D.M., Boucher, R.C., Randell, S.H., Suzuki, T., Gralinski, L.E., Kawaoka, Y., Baric, R.S., 2020. SARS-CoV-2 D614G variant exhibits efficient replication *ex vivo* and transmission *in vivo*. *Science*.
- Zhou, B., Thao, T.T.N., Hoffmann, D., Taddeo, A., Ebert, N., Labrousseau, F., Pohlmann, A., King, J., Steiner, S., Kelly, J.N., Portmann, J., Halwe, N.J., Ulrich, L., Trüeb, B.S., Fan, X., Hoffmann, B., Wang, L.I., Thomann, L., Lin, X., Stalder, H., Pozzi, B., de Brot, S., Jiang, N., Cui, D., Hossain, J., Wilson, M.M., Keller, M.W., Stark, T.J., Barnes, J.R., Dijkman, R., Jores, J., Benarafa, C., Wentworth, D.E., Thiel, V., Beer, M., 2021. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* 592 (7852), 122–127.
- van Dorp, L., Richard, D., Tan, C.C.S., Shaw, L.P., Acman, M., Balloux, F., 2020b. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun* 11, 5986.
- E. Volz, V. Hill, J.T. McCrone, A. Price, D. Jorgensen, A. O’Toole, J. Southgate, R. Johnson, B. Jackson, F.F. Nascimento, S.M. Rey, S.M. Nicholls, R.M. Colquhoun, A. da Silva Filipe, J. Shepherd, D.J. Pascall, R. Shah, N. Jesudason, K. Li, R. Jarrett, N. Pacchiari, M. Bull, L. Geidelberg, I. Siveroni, C.-U. Consortium, I. Goodfellow, N.J. Loman, O.G. Pybus, D.L. Robertson, E.C. Thomson, A. Rambaut, T.R. Connor, Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity, *Cell*, 184 (2021) 64-75 e11.
- Hillen, H.S., Kocic, G., Farnung, L., Dienemann, C., Tegunov, D., Cramer, P., 2020. Structure of replicating SARS-CoV-2 polymerase. *Nature* 584, 154–156.
- Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storic, 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.
- Ge, Y., Sun, S., 2020. Estimation of Coronavirus Disease Case-Fatality Risk in Real Time. *Emerg Infect Dis* 26, 1922–1923.
- Ioannidis, J.P.A., 2020. Infection fatality rate of COVID-19 inferred from seroprevalence data. *Bulletin of the World Health Organization*.
- 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).
- Young, B.E., Fong, S.W., Chan, Y.H., Mak, T.M., Ang, L.W., Anderson, D.E., Lee, C.Y., Amrun, S.N., Lee, B., Goh, Y.S., Su, Y.C.F., Wei, W.E., Kalimuddin, S., Chai, L.Y.A., Pada, S., Tan, S.Y., Sun, L., Parthasarathy, P., Chen, Y.Y.C., Barkham, T., Lin, R.T.P., Maurer-Stroh, S., Leo, Y.S., Wang, L.F., Renia, L., Lee, V.J., Smith, G.J.D., Lye, D.C., Ng, L.F.P., 2020. 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, 603–611.
- 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, I. Zhang, R.F. Jarrett, W.G. Glass, M. Beltramello, K. Nomikou, M. Pizzuto, L. Tong, E. Cameroni, T.I. Croll, N. Johnson, J. Di Iulio, A. Wickenhagen, A. Ceschi, A.M. Harbison, D. Mair, P. Ferrari, K. Smollett, F. Sallusto, S. Carmichael, C. Garzoni, J. Nichols, M. Galli, J. Hughes, A. Riva, A. Ho, M. Schiuma, M.G. Semple, P.J.M. Openshaw, E. Fadda, J.K. Baillie, J.D. Chodera, I.C. Investigators, C.-G.U. Consortium, S.J. Rihn, S.J. Lycett, H.W. Virgin, A. Telenti, D. Corti, D.L. Robertson, G. Snell, Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity, *Cell*, 184 (2021) 1171-1187 e1120.
- Adlhoj, C., Mook, P., Lamb, F., Ferland, L., Melidou, A., Amato-Gauci, A.J., Pebody, R. N., 2021. European Influenza Surveillance, Very little influenza in the WHO European Region. *Euro Surveill* 26 during the 2020/21 season, weeks 40 2020 to 8 2021.
- Uyeki, T.M., Santoli, J., Jernigan, D.B., 2020. Preparing for the 2020–2021 Influenza Season. *JAMA* 324, 2318–2319.
- WHO Global Influenza Surveillance and Response System (GISRS). Number of Specimens Positive for Influenza by Subtype, Eastern Mediterranean Region of WHO (2021) <https://apps.who.int/flumart/Default?ReportNo=3&WHORegion=> Last Accessed: Mar 21 2021.
- Jones, N., 2020. How COVID-19 is changing the cold and flu season. *Nature* 588, 388–390.
- Kim, D., Quinn, J., Pinsky, B., Shah, N.H., Brown, I., 2020. Rates of Co-infection Between SARS-CoV-2 and Other Respiratory Pathogens. *JAMA* 323, 2085–2086.
- Kim, K.W., Deveson, I.W., Pang, C.N.I., Yeang, M., Naing, Z., Adikari, T., Hammond, J. M., Stevanovski, I., Beukers, A.G., Verich, A., Yin, S., McFarlane, D., Wilkins, M.R., Stelzer-Braid, S., Bull, R.A., Craig, M.E., van Hal, S.J., Rawlinson, W.D., 2021. Respiratory viral co-infections among SARS-CoV-2 cases confirmed by virome capture sequencing. *Sci Rep* 11, 3934.
- A. Elghoudi, H. Aldhanhani, G. Ghatasheh, E. Sharif, H. Narchi, Covid-19 in Children and Young Adolescents in Al Ain, United Arab Emirates- a Retrospective Cross-Sectional Study, *Front Pediatr*, 8 (2020) 603741.
- Saddique, A., Rana, M.S., Alam, M.M., Ikram, A., Usman, M., Salman, M., Faryal, R., Massab, U., Bokhari, H., Mian, M.S., Israr, A., 2020. Safiullah, Emergence of co-infection of COVID-19 and dengue: A serious public health threat. *J Infect* 81, e16–e18.
- Lansbury, L., Lim, B., Baskaran, V., Lim, W.S., 2020. Co-infections in people with COVID-19: a systematic review and meta-analysis. *J Infect* 81, 266–275.
- Keaten, J., 2020. WHO: 10% of world’s people might have had COVID-19, *Telegraph - Herald. Dubuque, Iowa*. <https://apnews.com/article/virus-outbreak-united-nation-s-health-ap-top-news-international-news-54a3a5869c9ae4ee623497691e796083>. Last accessed: 19 July 2021.