








Article

Correlates of SARS-CoV-2 Variants on Deaths, Case Incidence and Case Fatality Ratio among the Continents for the Period of 1 December 2020 to 15 March 2021

Wajdy J. Al-Awaida ^{1,*}, Baker Jawabrah Al Hourani ², Samer Swedan ³, Refat Nimer ³, Foad Alzoughool ⁴, Hamzeh J. Al-Ameer ¹, Sara E. Al Tamam ¹, Raghad Alashqar ¹, Omar Al bawareed ⁵, Yulia Gushchina ⁶, Mohamed Samy Abousenna ⁷, Amani Marwan Ayyash ⁸, Ahmad Sharab ¹, Sulaiman M. Alnaimat ⁹, Amer Imraish ¹⁰, Muhanad W. Akash ¹¹ and Najah R. Hadi ¹²

- ¹ Department of Biology and Biotechnology, Faculty of Science, American University of Madaba, Madaba 11821, Jordan; h.alameer@aum.edu.jo (H.J.A.-A.); saratt2001148@gmail.com (S.E.A.T.); 1910332@std.aum.edu.jo (R.A.); a.sharab@aum.edu.jo (A.S.)
 - ² Department of Biology and Chemistry, Embry-Riddle Aeronautical University, Prescott, AZ 86301, USA; Jawabrab@erau.edu
 - ³ Department of Medical Laboratory Sciences, Jordan University of Science and Technology, Irbid 22110, Jordan; sfswedan4@just.edu.jo (S.S.); rnmimer@just.edu.jo (R.N.)
 - ⁴ Department of Medical Laboratory Sciences, The Hashemite University, Zarqa 13133, Jordan; foad@hu.edu.jo
 - ⁵ Department of Normal Physiology, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Street, 117198 Moscow, Russia; alomar442@mail.ru
 - ⁶ Department of General and Clinical Pharmacology, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Street, 117198 Moscow, Russia; gushchina-yush@rudn.ru
 - ⁷ Central Laboratory for Evaluation of Veterinary Biologics, Agricultural Research Center, Cairo 11517, Egypt; mohamedsamy2020@hotmail.com
 - ⁸ Department of Pharmacy, Faculty of Health Sciences, American University of Madaba, Madaba 11821, Jordan; a.ayyash@aum.edu.jo
 - ⁹ Biological Department, Faculty of Science, Al Hussein Bin Talal University, Ma'an P.O. Box 20, Jordan; s_alnaimat@ahu.edu.jo
 - ¹⁰ Department of Biology, Faculty of Science, The University of Jordan, Amman 11942, Jordan; a.imraish@ju.edu.jo
 - ¹¹ Department of Horticulture and Crop Science, Faculty of Agriculture, The University of Jordan, Amman 11942, Jordan; makash@ju.edu.jo
 - ¹² Department of Pharmacology and Therapeutics, Faculty of Medicine, University of Kufa, Najaf 540011, Iraq; drnajahhadi@yahoo.com
- * Correspondence: w.alawaida@aum.edu.jo; Tel.: +962-5-3294444 (ext. 5614)



Citation: J. Al-Awaida, W.; Jawabrah Al Hourani, B.; Swedan, S.; Nimer, R.; Alzoughool, F.; J. Al-Ameer, H.; E. Al Tamam, S.; Alashqar, R.; Al bawareed, O.; Gushchina, Y.; et al. Correlates of SARS-CoV-2 Variants on Deaths, Case Incidence and Case Fatality Ratio among the Continents for the Period of 1 December 2020 to 15 March 2021. *Genes* **2021**, *12*, 1061. <https://doi.org/10.3390/genes12071061>

Academic Editors: Paolo Cinelli and Michele Rubini

Received: 17 May 2021

Accepted: 9 July 2021

Published: 12 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The outbreak of coronavirus disease 2019 (COVID-19), by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has quickly developed into a worldwide pandemic. Mutations in the SARS-CoV-2 genome may affect various aspects of the disease including fatality ratio. In this study, 553,518 SARS-CoV-2 genome sequences isolated from patients from continents for the period 1 December 2020 to 15 March 2021 were comprehensively analyzed and a total of 82 mutations were identified concerning the reference sequence. In addition, associations between the mutations and the case fatality ratio (CFR), cases per million and deaths per million, were examined. The mutations having the highest frequencies among different continents were Spike_D614G and NSP12_P323L. Among the identified mutations, NSP2_T153M, NSP14_I42V and Spike_L18F mutations showed a positive correlation to CFR. While the NSP13_Y541C, NSP3_T73I and NSP3_Q180H mutations demonstrated a negative correlation to CFR. The Spike_D614G and NSP12_P323L mutations showed a positive correlation to deaths per million. The NSP3_T1198K, NS8_L84S and NSP12_A97V mutations showed a significant negative correlation to deaths per million. The NSP12_P323L and Spike_D614G mutations showed a positive correlation to the number of cases per million. In contrast, NS8_L84S and NSP12_A97V mutations showed a negative correlation to the number of cases per million. In addition, among the identified clades, none showed a significant correlation to CFR. The G, GR, GV, S clades showed a significant positive correlation to deaths per million. The GR and S clades showed a positive correlation to number of cases per million. The clades having the highest

frequencies among continents were G, followed by GH and GR. These findings should be taken into consideration during epidemiological surveys of the virus and vaccine development.

Keywords: SARS-CoV-2; COVID-19; GISAID; deaths per million; cases per million; Spearman's correlation; genomic variations; case fatality ratio

1. Introduction

The novel betacoronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), responsible for coronavirus disease 2019 (COVID-19), was first described in Wuhan, China in December 2019 [1–3]. Subsequently, COVID-19 became a worldwide pandemic. According to the World Health Organization (WHO), for the period of 1 December 2020 to 15 March 2021, the worldwide outbreak of COVID-19 has approximately 64,668,248 cases and more than 1,339,685 deaths were reported. SARS-CoV-2 is an enveloped single-stranded positive RNA virus [4]. The SARS-CoV-2 complete genome consists of about 30,000 nucleotides that translate structural and non-structural proteins (Nsps) [1,2]. Coronaviruses share a roughly spherical or moderately pleomorphic shape with distinct spike (S) protein envelope projections [5]. The four SARS-CoV-2 structural proteins are the spike (S), the envelope (E), the membrane (M) and the nucleocapsid (N) proteins, while nonstructural proteins are 3-chymotrypsin-like protease (nsp3 and nsp5), papain-like protease and RNA-dependent RNA polymerase (nsp12) [6,7]. Surface S glycoprotein attaches to the host's angiotensin-converting enzyme 2 (ACE2) receptor and plays an important role in viral adsorption and human transmission [8]. Nsps, created as cleavage products from open reading frame 1ab (ORF1ab), are assembled to facilitate transcription and viral replication. Viral RNA is synthesized and controlled by the RNA-dependent polymerase with assistance from Nsp7 and Nsp88 [9].

SARS-CoV-2 genetic mutations may allow the virus to adapt to a new host cell or alter virus pathogenesis [10]. Therefore, detecting and analyzing virus mutations in different countries could enhance our understanding of the contribution of such mutations to viral pathogenesis and spread. In contrast to SARS-CoV in 2002 and Middle East Coronavirus Respiratory Syndrome (MERS-CoV) in 2012, SARS-CoV-2 spread worldwide quickly. To date, the greatest assessment of the fatality ratio of SARS-CoV-2 has come from data recorded in China, France and the Diamond Princess cruise ship [11]. Effective vaccines and treatments are urgently needed to decrease the mortality and morbidity ratio due to COVID-19. Variations have been observed in mortality ratios from country to country [3]. However, factors explaining these variations have not been elucidated. Furthermore, the contribution of SARS-CoV-2 genetic variations and host genetic factors to mortality ratio remains unknown.

In this study, we evaluated 553,518 SARS-CoV-2 genome sequences isolated from different continents and explored their associations with CFR, cases per million and deaths per million.

2. Materials and Methods

2.1. Data Availability

Full-length viral nucleotide sequence of the reference SARS-CoV-2 Wuhan-Hu-1 strain (Accession NC_045512) [1] was obtained from the global initiative on sharing all influenza data (GISAID). We used a total of 553,518 SARS-CoV-2 genome sequences isolated from patients/individuals in different continents, including 157,691 from North America, 6702 from South America, 346,951 from Europe, 1670 from Australia, 34,736 from Asia and 5768 from Africa. SARS-CoV-2 genome sequences were collected from a publicly open database, the GISAID (<https://www.gisaid.org/> (accessed on 1 December 2020)).

2.2. Mutation Analysis

Only high-coverage SARS-CoV-2 sequences were downloaded from GISAID Epi-CoVTM database [12]. Each of the SARS-CoV-2 sequences was compared with the reference sequence SARS-CoV-2 Wuhan-Hu-1 (accession number MN908947.3) by using CoVsurver enabled by GISAID (<https://www.gisaid.org/epiflu-applications/covsurvermutations-app/> (accessed on 1 December 2020)) [12,13]. After the alignment of the full-length viral nucleotide sequence, we identified SARS-CoV-2 mutations in each of the isolated virus sequences. Single mutations based on a population average were used for the estimation of the CFR.

Genomic clades were inferred by GISAID database and defined according to its nomenclature system at the time of data collection outlined in (<https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/> (accessed on 1 December 2020)).

2.3. Data Acquisition

The number of deaths and confirmed cases related to COVID-19 was collected from the Worldometer (<https://www.worldometers.info/coronavirus/> (accessed on 1 December 2020)). The CFR was calculated by dividing total deaths by total confirmed cases in each country/region.

2.4. Phylogenetic Analysis

In comparison with the reference sequence of SARS-CoV-2 Wuhan-Hu-1 strain (Accession NC_045512.2), phylogenetic analysis was performed for the complete coding region of all isolates using Nextstrain software [14].

2.5. Ethical Statement

The study was carried out following the Helsinki Declaration. Ethical approval was obtained from the Institutional Review Board at the American University of Madaba (Ref. H20010).

2.6. Statistical Analyses

The data were analyzed using the Social Sciences Statistical Package version 23 (IBM Inc., Armonk, New York, NY, USA). Hierarchical cluster analysis was used to identify mutant clusters among the countries/regions. Spearman's correlation (r) values and two-tailed significance (p) were used to identify the correlation between mutation frequencies with CFR, cases per million and deaths per million. A p -value of less than 0.05 was considered to be statistically significant.

3. Results

COVID-19 related information for different continents is shown in Table 1. Africa had the highest CFR of 2.56%; Australia had the lowest CFR of 0.79% (Table 1). The highest recorded number of cases per million in Europe (58,437), while the lowest was in Africa (6295). The highest recorded number of deaths per million was in Europe (1117), while the lowest was in Australia (45).

3.1. Clades Frequencies of SARS-CoV-2 Genome

The clades for 553,518 SARS-CoV-2 genome sequences from different world continents were identified. The clades frequencies are shown in Figure 1A. The clades having the highest frequencies among continents were G, GH and GR. The highest frequencies for G clade were in South America and the predominance of clade GH in North America. The clades GR predominated in Asia (Supplementary Table S1).

Table 1. COVID-19 statistics in the different world continents.

Region	Population	Total Cases	Cases/1M	Total Deaths	Deaths/1M	Case Fatality Ratio	Virus Isolates
Africa	1,363,047,821	3,067,758	6295	78,435	101	2.56	5768
Asia	4,624,453,127	12,546,291	17,067	169,912	177	1.35	34,736
Australia	42,643,478	24,777	7705	196	45	0.79	1670
Europe	747,950,109	23,239,185	58,437	532,996	1117	2.29	346,951
North America	592,362,766	17,757,444	22,816	380,088	333	2.14	157,691
South America	433,264,918	8,032,793	30,293	178,058	701	2.22	6702

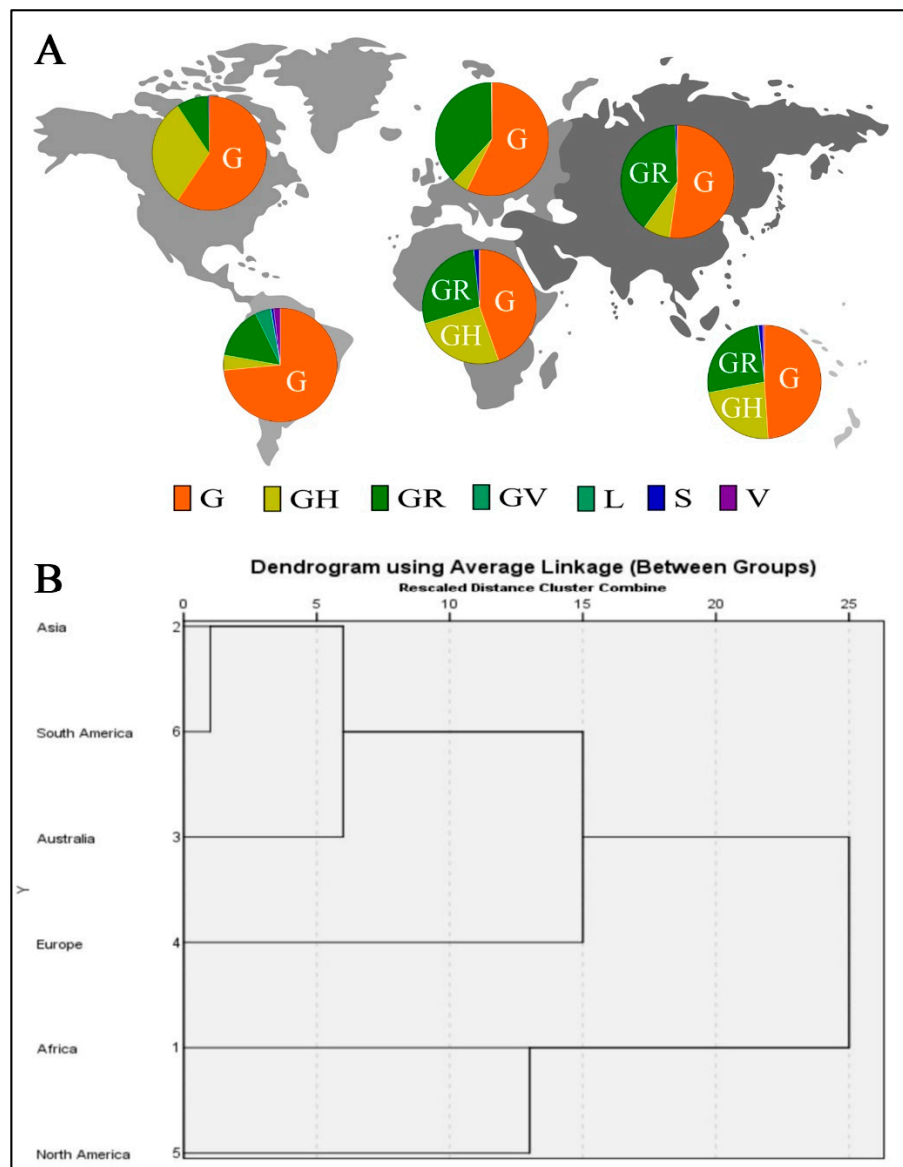


Figure 1. (A) Continent distribution of various SARS-CoV-2 clades. (B) Dendrogram showing clustering of SARS-CoV-2 isolates from different continents.

The cluster analysis of the mutations indicates similarity among SARS-CoV-2 isolates from South America and Asia; isolates from Australia have a similarity to those in America and Asia and a lesser degree similarity among SARS-CoV-2 isolates from Europe and Africa. On the other hand, isolates from North America were least similar to those from other regions (Figure 1B).

3.2. Correlation between SARS-CoV-2 Genetic Clades with Case Fatality Ratio (CFR), Cases per Million and Deaths per Million

Table 2 demonstrates Spearman's correlation (r) values and two-tailed significance (p) values between the SARS-CoV-2 clades and each of the CFR, cases per million and deaths per million. Among the identified clades, none showed a significant correlation to CFR. The G, GR, GV, S clades showed a significant positive correlation to deaths per million. The GR and S clades showed a significant positive correlation to number of cases per million.

Table 2. Correlation analysis between SARS-CoV-2 clades and Case fatality rates, cases per million and deaths per million among the identified clades in all world continents.

Clade	Case Fatality Ratio		Deaths/1M		Cases/1M	
	r	p	r	p	r	p
Clade G	0.257	0.623	0.943	0.005 **	1	-
Clade GH	0.086	0.872	0.6	0.2	0.771	0.072
Clade GR	0.371	0.468	0.829	0.042 *	0.943	0.005 **
Clade GV	0.371	0.468	0.829	0.042 *	0.657	0.156
Clade GL	0.086	0.872	0.771	0.072	0.6	0.208
Clade S	0.371	0.468	0.829	0.042 *	0.943	0.005 **
Clade V	0.086	0.872	0.771	0.072	0.6	0.208

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed). r : Spearman's correlation. p : p -value.

3.3. Correlation between SARS-CoV-2 Genetic Variants with Case Fatality Ratio (CFR), Cases per Million and Deaths per Million

Supplementary Table S2 demonstrates Spearman's correlation (r) values and two-tailed significance (p) values between the SARS-CoV-2 mutations and each of the CFR, cases per million and deaths per million. Among the identified mutations, NSP2_T153M, NSP14_I42V and Spike_L18F mutations showed a positive correlation to CFR (Figure 2A–C). However, NSP3_T73I and NSP3_Q180H, NSP13_Y541C showed a negative correlation to CFR (Figure 3A–C). Spike_D614G and NSP12_P323L mutations displayed a positive correlation to deaths per million (Figure 4A,B). The NSP3_T1198K, NS8_L84S and NSP12_A97V mutations showed a significant negative correlation to deaths per million (Figure 4C–E). NSP12_P323L and Spike_D614G mutations showed a positive correlation to number of cases per million (Figure 5A,B). However, NS8_L84S and NSP12_A97V mutations showed a negative correlation to the number of Cases per million (Figure 5C,D).

3.4. Global Mapping of Frequencies of the Mutations Correlating with CFR, Cases per Million and Deaths per Million among World Continents

Supplementary Table S3 demonstrates the clustering of SARS-CoV-2 isolates and heatmap of the frequency of mutations of the SARS-CoV-2 genome by continent. The NSP2_T153M mutation was reported at 1.18%, 0.02%, 0.04%, 0.07% and 0.07% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Europe, North America and South America, respectively (Figure 2D). The NSP14_I42V mutation was reported at 0.017 and 0.003 in SARS-CoV-2 genome sequences isolated from Africa and Europe, respectively (Figure 2E). The Spike_L18F mutation was reported at 16.02%, 0.72%, 1.86%, 10.90%, 1.25% and 5.94% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe, North America and South America, respectively (Figure 2F).

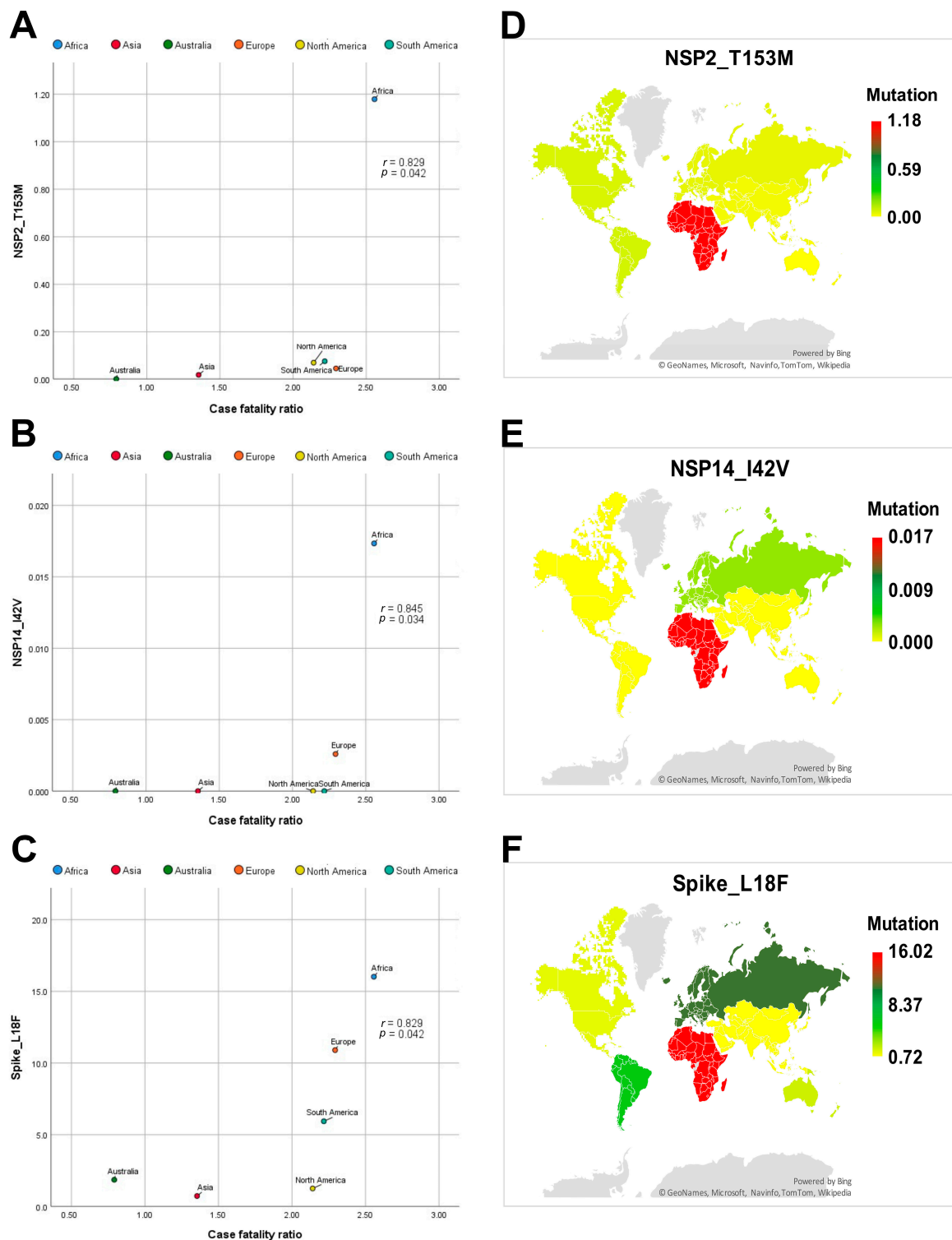


Figure 2. Correlation analysis and global mapping of mutations having a positive correlation with CFR among different world continents. (A) Spearman's correlation between NSP2_T153M mutation and CFR among different world continents, (B) Spearman's correlation between NSP14_I42V mutation and CFR among different world continents, (C) Spearman's correlation between Spike_L18F mutation and CFR among different world continents, (D) Global mapping of NSP2_T153M mutation having a positive correlation with CFR among different world continents. (E) Global mapping of NSP14_I42V mutation having a positive correlation with CFR among different world continents. (F) Global mapping of Spike_L18F mutation having a positive correlation with CFR among different world continents: r : Spearman's correlation coefficient. p : p -value.

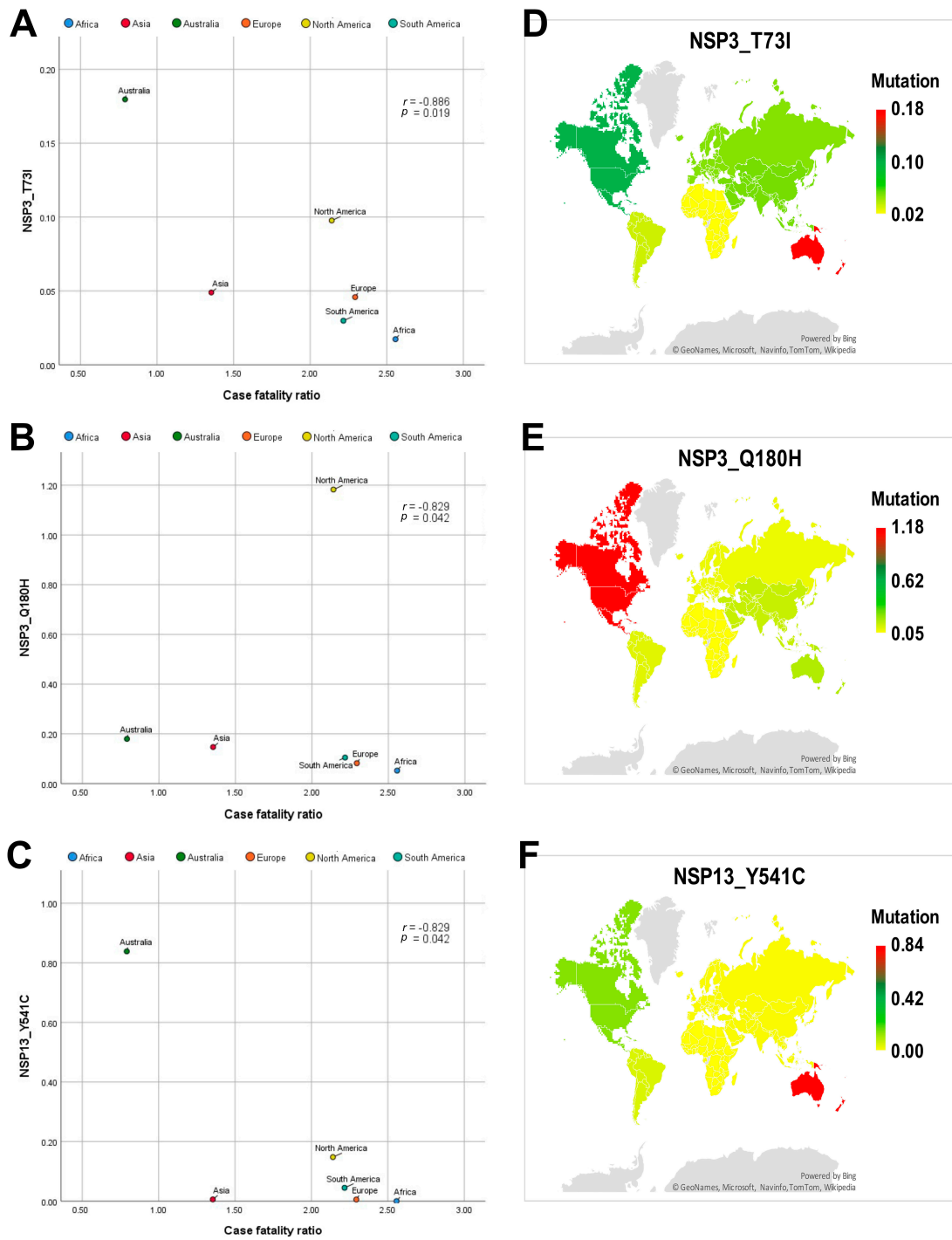


Figure 3. Correlation analysis and global mapping of mutations having a negative correlation with CFR among different world continents. (A) Spearman's correlation between NSP3_T73I mutation and CFR among different world continents, (B) Spearman's correlation between NSP3_Q180H mutation and CFR among different world continents, (C) Spearman's correlation between NSP13_Y541C mutation and CFR among different world continents, (D) Global mapping of NSP3_T73I mutation having a negative correlation with CFR among different world continents. (E) Global mapping of NSP3_Q180H mutation having a negative correlation with CFR among different world continents. (F) Global mapping of NSP13_Y541C mutation having a negative correlation with CFR among different world continents: r : Spearman's correlation coefficient. p : p -value.

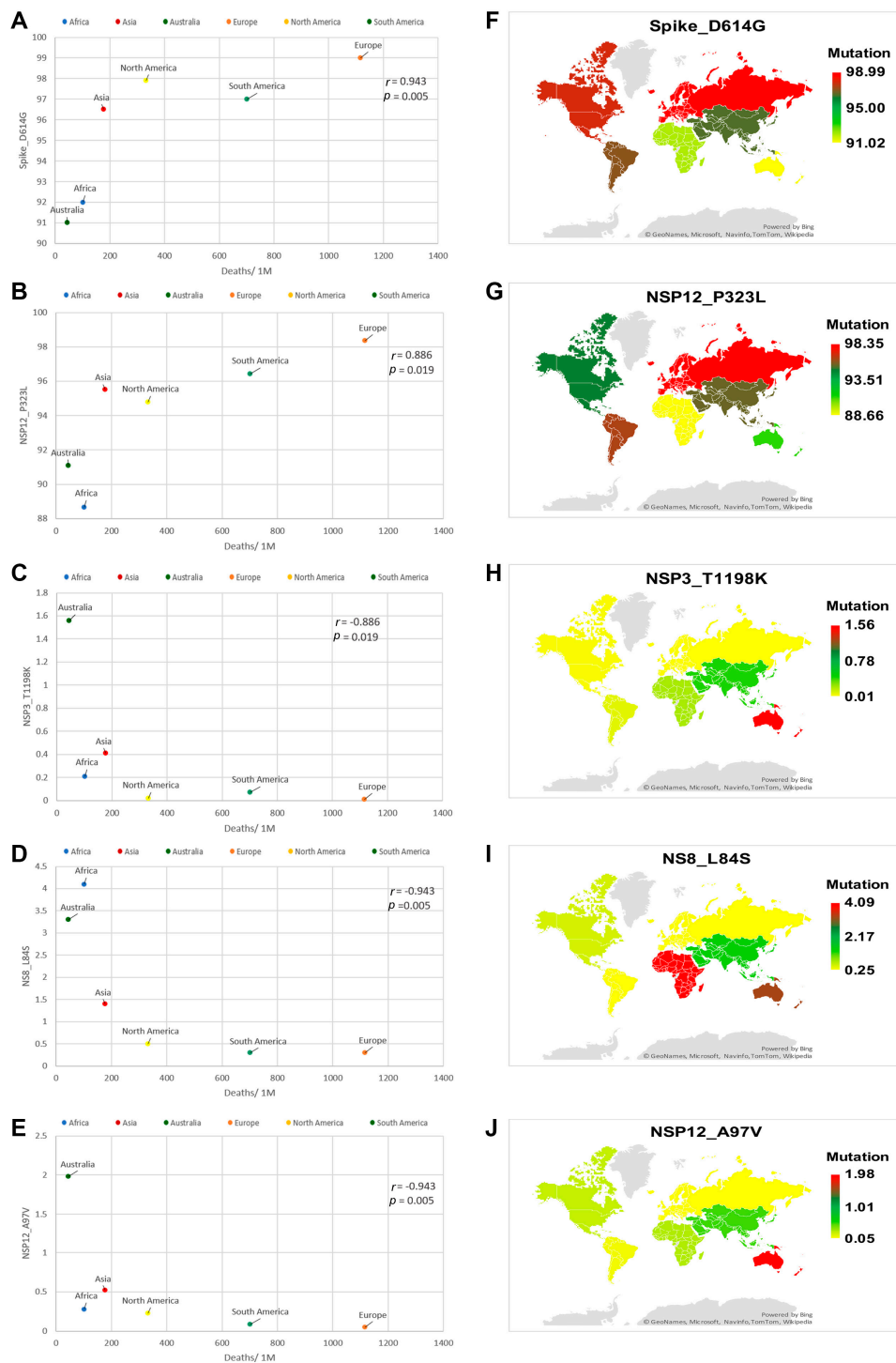


Figure 4. Correlation analysis and global mapping of mutations having a positive correlation with deaths per million among different world continents. (A) Spearman's correlation between Spike_D614G mutation and deaths per million, (B) Spearman's correlation between NSP12_P323L mutation and deaths per million, (C) Spearman's correlation between NSP3_T1198K mutation and deaths per million. (D) Spearman's correlation between NS8_L84S mutation and deaths per million. (E) Global mapping of NSP12_A97V mutation having a positive correlation with deaths per million. (F) Global mapping of Spike_D614G mutation having a positive correlation with deaths per million. (G) Global mapping of NSP12_P323L mutation having a positive correlation with deaths per million. (H) Global mapping of NSP3_T1198K mutation having a negative correlation with deaths per million. (I) Global mapping of NS8_L84S mutation having a negative correlation with deaths per million. (J) Global mapping of NSP12_A97V mutation having a negative correlation with deaths per million r : Spearman's correlation coefficient. p : p -value.

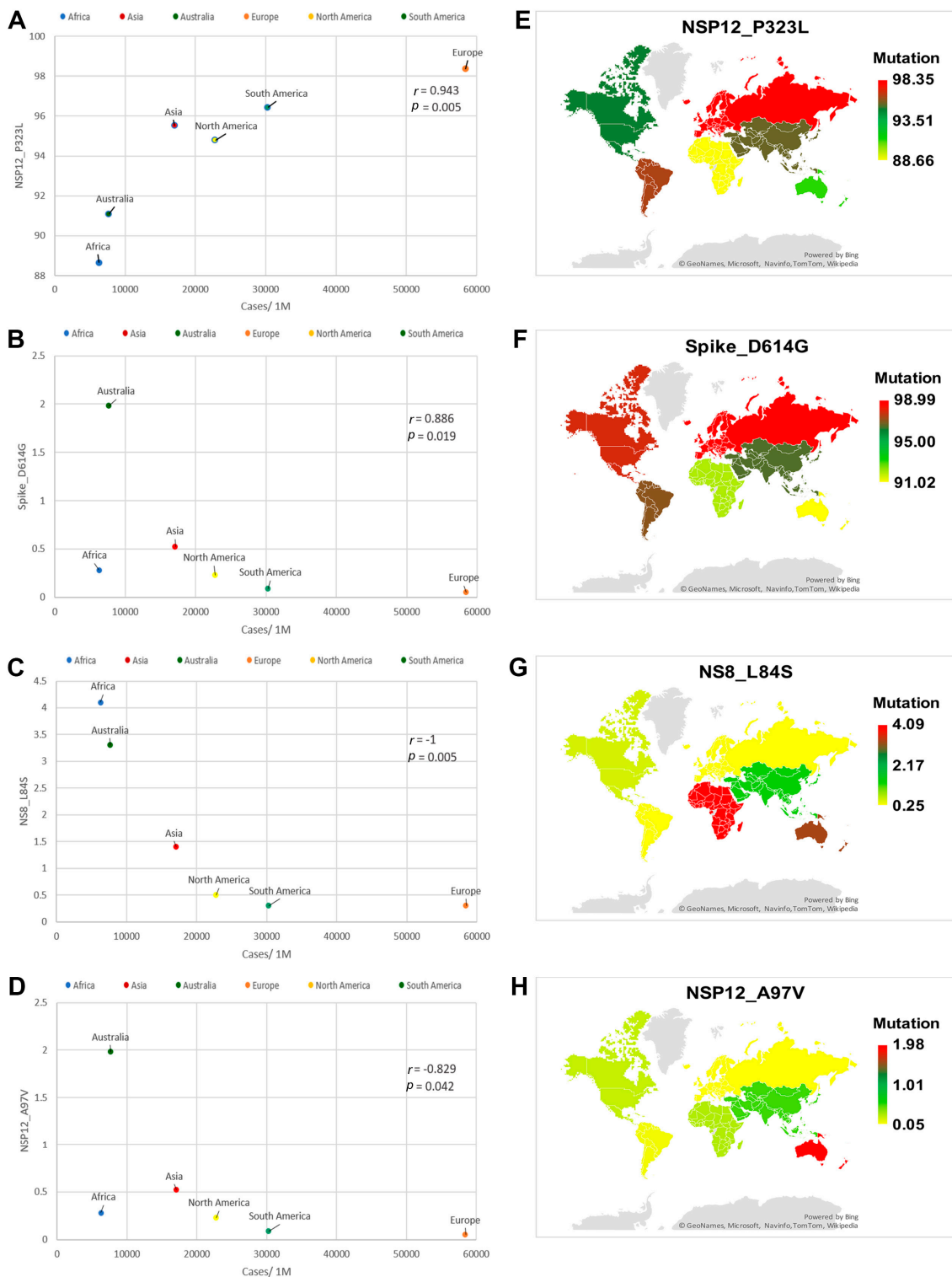


Figure 5. Correlation analysis and global mapping of mutations with cases per million among different world continents. (A) Spearman's correlation between NSP12_P323L mutation and cases per million, (B) Spearman's correlation between Spike_D614G mutation and cases per million (C) Spearman's correlation between NS8_L84S mutation and cases per million. (D) Spearman's correlation between NSP12_A97V mutation and cases per million. (E) Global mapping of NSP12_P323L mutation having a positive correlation with cases per million. (F) Global mapping of Spike_D614G mutation having a positive correlation with cases per million. (G) Global mapping of NS8_L84S mutation having a negative correlation with cases per million. (H) Global mapping of NSP12_A97V mutation having a negative correlation with cases per million. r : Spearman's correlation coefficient. p : p -value.

The NSP3_T73I mutation was reported at 0.02%, 0.05%, 0.18%, 0.05%, 0.10% and 0.03% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe, North America and South America, respectively (Figure 3D). The NSP3_Q180H mutation was reported at 0.05%, 0.15%, 0.18%, 0.08%, 1.18% and 0.10% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe, North America and South America, respectively (Figure 3E). The NSP13_Y541C mutation was reported at 0.01%, 0.84%, 0.01%, 0.15% and 0.04% in SARS-CoV-2 genome sequences isolated from Asia, Australia, Europe, North America and South America, respectively (Figure 3F).

The Spike_D614G mutation was reported at 91.97%, 96.46%, 91.02%, 98.99%, 97.91% and 97.02% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe, North America and South America, respectively (Figures 4F and 5F). The NSP12_P323L mutation was reported at 88.66%, 95.52%, 91.08%, 98.35%, 94.79% and 96.43% in the SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe, North America and South America, respectively (Figures 4G and 5E). The NSP3_T1198K mutation was reported at 0.21%, 1.56%, 0.18%, 0.01%, 0.02% and 0.07% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe, North America and South America, respectively (Figure 4H). The NS8_L84S mutation was reported at 4.09%, 1.39%, 3.29%, 0.25%, 0.51% and 0.25% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe, North America and South America, respectively (Figures 4I and 5G). The NSP12_A97V mutation was reported at 0.28%, 0.52%, 1.98%, 0.05%, 0.23% and 0.09% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe, North America and South America, respectively (Figures 4J and 5H).

Other mutations have a frequency higher than 20% without having a significant correlation with CFR, cases per million and deaths per million among world continents. The N_G204R mutation was reported at 22.31%, 75.69%, 46.59%, 52.36% and 76.01% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe and South America, respectively. The N_R203K mutation was reported at 24.24%, 75.88%, 47.90%, 53.51%, 15.08% and 76.41% in SARS-CoV-2 genome sequences isolated from Africa, Asia, Australia, Europe, North America and South America, respectively. The NSP2_T85I mutation was reported at 40.20% and 59.59% in SARS-CoV-2 genome sequences isolated from Africa and South America, respectively. The NS3_Q57H mutation was reported at 46.10%, 32.99% and 63.30% in SARS-CoV-2 genome sequences isolated from Africa, Australia and North America, respectively. The Spike_E484K mutation was reported at 35.63% in SARS-CoV-2 genome sequences isolated from Africa. The Spike_N501Y mutation was reported at 36.55% and 47.23% in SARS-CoV-2 genome sequences isolated from Africa and Europe, respectively. The Spike_H69del mutation was reported at 48.93% in SARS-CoV-2 genome sequences isolated from Europe. The Spike_K417N mutation was reported at 34.36% in SARS-CoV-2 genome sequences isolated from Africa. The NS8_S24L mutation was reported at 33.88% in SARS-CoV-2 genome sequences isolated from North America. The NS8_Q27stop mutation was reported at 46.36% in SARS-CoV-2 genome sequences isolated from Europe. The Spike_A222V mutation was reported at 26.59% in SARS-CoV-2 genome sequences isolated from Europe. The Spike_V1176F mutation was reported at 29.30% in SARS-CoV-2 genome sequences isolated from South America. The N_I292T mutation was reported at 20.50% in SARS-CoV-2 genome sequences isolated from South America. The NS3_S171L mutation was reported at 30.37% in SARS-CoV-2 genome sequences isolated from Africa.

3.5. Phylogenetic Clades Tree of SARS-CoV-2 in Different World Continents

The phylogenetic analysis indicated that several clades were present on different continents. The analysis also revealed that clades 20 became dominant. Clades 20 (including 20A, 20B, 20C, 20E (EU1), 20F, 20G, 20H/501Y.V2, 20I/501Y.V1 and 20J/501Y.V2) by Nextstrain nomenclature, are also known as clades G, GR and GH by GISAID nomenclature (Figure 6).

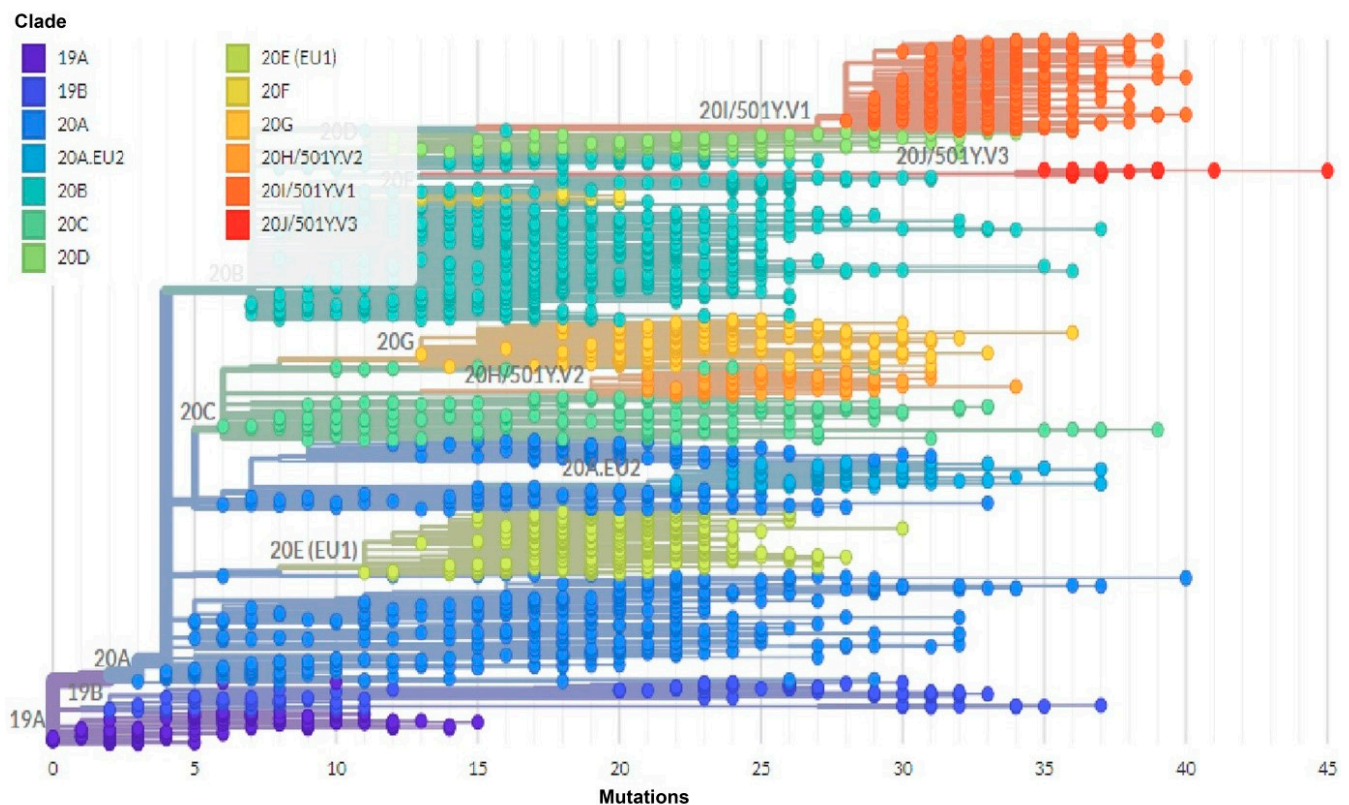


Figure 6. Phylogenetic analysis of SARS-CoV-2 isolates from different continents. SARS-CoV-2 clades were present in different continents.

4. Discussion

The current COVID-19 outbreak has spread rapidly worldwide. Most COVID-19 infected patients suffering from COVID-19 are either asymptomatic or experience mild symptoms. Around 15% of all COVID-19 patients progress to severe pneumonia and approximately 5% of COVID-19 patients have acute respiratory disorder syndrome, septic shock and multiple organ failure [15,16]. The CFR associated with COVID-19 differs from continent to continent and is higher in Africa, Europe, South America and North America countries compared with those in Asia and Australia. While several hypotheses have been put forward to explain the variations in CFR, including age distribution differences, blood groups, gender, virus genomic types and ethnic backgrounds, many of these hypotheses still require confirmation.

This study explored and compared mutation profiles of SARS-CoV-2 isolates from different continents. We identified differences among continents relating to COVID-19 CFR, cases per million and deaths per million. The differences may be attributed to differences in age distributions [17,18], blood groups [19], gender [20,21], virus genomic types [3], strict lockdown strategies [22] social distancing [23] and genetic backgrounds [24,25]. Hence, this study also investigated the contribution of genetic variations in the SARS-CoV-2 genome to the COVID-19 CFR, the number of cases per million and the number of deaths per million.

Like other RNA viruses, genetic variation in SARS-CoV-2 is important for fitness, survival and pathogenesis. Spontaneous mutations and recombination are two main sources of genetic variation for SARS-CoV-2 [26].

Based on data obtained from a public database GISAID for the period 1 December 2020 to 15 March 2021, the mutation rate of the SARS-CoV-2 virus is about 8×10^{-4} nucleotides/genome per year [27], which is considered a high rate for an RNA virus [27–29]. Analysis of 553,518 genomic sequences from the GISAID databases showed that the number and the frequency of mutations in Africa and Europe were considerably higher than in Asia [30].

The identification of SARS-CoV-2 mutation patterns indicated differences relating to geography, time of isolation and subject age, but not relating to gender [31].

The mutation is one of the most effective evolutionary mechanisms in RNA viruses [32]. SARS-CoV-2 genomic studies described mutations in several genes, including ORF1ab, ORF3a N, S, M, E, ORF8, ORF7, ORF6 and ORF10 [33–35]. The S, ORF1ab, NSP15, NSP12, NSP3, NSP2, NSP1 and ORF8 genes had significantly higher mutation frequencies than other genes [33,36,37]. Several studies have been carried out to identify genomic variants of SARS-CoV-2, including synonymous, nonsynonymous, deletion, addition and non-coding mutations [31,33,38]. The most common mutations among the SARS-CoV-2 genome were nonsynonymous and synonymous mutations [31,36]. In the current study, the mutations having the highest frequencies among continents were missense, including Spike_D614G and NSP12_P323L. Moreover, the highest frequencies for Spike_D614G and NSP12_P323L mutations were among isolates from Europe with 98.99% and 98.35%, respectively.

Our study indicated a significant correlation between the number of both the number of cases per million and the number of deaths per million, with the Spike_D614G mutation, which is a mutation outside the receptor-binding domain (RBD). In line with previous studies, this study revealed that the Spike_D614G mutation showed a positive correlation with viral infectivity and enhanced transmissibility [39–41].

The genome of SARS-CoV-2 is about 30 kb and codes for 14 open frames (ORFs). 5'-ORF-1a/1b forms 67% of the genome and encodes polyproteins 1a (pp1a) and polyproteins 1 ab (pp1ab), which are then processed into sixteen non-structural proteins (nsp1-16) [42]. Other nsp proteins are believed to be associated with viral replication or host immune response modulation [43]. In our study, many genetic variants in non-structural proteins, including NSP3_T73I, NSP3_Q180H and NSP13_Y541C mutations showed a significant negative correlation to CFR. In addition, NSP3_T1198K, NS8_L84S and NSP12_A97V mutations showed a significant negative correlation to the deaths per million. NSP12_P323L mutation showed a positive correlation to deaths per million. The NS8_L84S and NSP12_A97V mutations showed a negative correlation to number of cases per million. In addition, the NSP1_M85I, NSP3_E475D showed a significant positive correlation to the number of cases per million. These mutations may affect viral replication or host immune response modulation, leading to change in viral pathogenicity [44].

The trimetric viral spike (S) glycoprotein is the main determinant of coronaviruses' host specificity and pathogenesis. The S2 domain activates the viral and host cell membranes' fusion to encourage viral penetration and uncoating [45,46]. In our study, The Spike_L18F mutation showed a significant positive correlation to CFR. On the other hand, The Spike_D614G mutation showed a significant positive correlation to the number of cases per million and the number of deaths per million. According to our findings, The Spike_D614G mutation in the S2 domain has a positive correlation between viral infectivity and transmissibility.

A recent report have shown that the crystal structure of the SARS-CoV-2 nucleocapsid (N) protein is identical to that of the known coronavirus N proteins, but their surface electrostatic potential properties were distinct [47]. Sheik et al. examined the variables impacting the variation of the N gene among 13 coronaviruses, as well as how they affect the virus-host relationship, reporting high AT% and low GC% in the coronavirus SARS N gene [48]. The N protein of coronaviruses plays a major role in the packaging of the viral genome and virus assembly [49]. In our study, NS8_L84S mutation showed a significant negative correlation to deaths per. These mutations may have a critical role in multiple steps of the viral life cycle.

The phylogenetic analysis indicated that several virus clades were deployed in different continents and it was observed that clades 20 became dominant. Clades 20 included 20A, 20B, 20C, 20E(EU1), 20F, 20G, 20H/501Y.V2, 20I/501Y.V1 and 20J/501Y.V2 by Nextstrain nomenclature, which is also known as clades G, GR and GH by GISAID nomenclature [50]. The majority of genomes were classified into one of seven major clades: L, S, V, G, GH, GR, or GV [51–53]. The clades G, GH, GR and GV, which account for 98% of the genomes, were

found to have the D614G mutation. Clade G was identified the most frequently, followed by GR and GH. The mutation was discovered in a glycosylated residue of the viral spike, which is conserved throughout this species [54]. Mutations in this area could be associated with alterations in host cell membrane fusion capacity [53], an effect that should also result in increased transmission and pathogenicity between individuals. Korber and colleagues subsequently provided experimental evidence that associate this mutation in COVID-19 patients with increased infectivity and greater viral loads [39]. Sub-clusters of clade G began to evolve into clades GH, GR and, more recently, clade GV. The current study's analysis of the chronological distribution of SARS-CoV-2 clades revealed a significant increase in the number of sequenced genomes clustered into the GR clade relative to clade G. Additionally, a change in the number of genomes clustered into clade GH was observed. Finally, on the basis of the details presented, the adaptation-driven genetic evolution hypothesis appears to be more viable. However, experimental evidence designed to allow for comparison of clades needs to be established.

5. Conclusions

The mutations having the highest frequencies among continents were Spike_D614G and NSP12_P323L. Among the identified mutations, NSP2_T153M, NSP14_I42V and Spike_L18F mutations showed positive correlation to CFR, while NSP13_Y541C, NSP3_T73I and NSP3_Q180H showed a negative correlation. The Spike_D614G and NSP12_P323L mutations showed positive correlation to deaths per million, while NSP3_T1198K, NS8_L84S and NSP12_A97V mutations showed negative correlation. NSP12_P323L and Spike_D614G mutations showed positive correlation to number of cases per million, while NS8_L84S and NSP12_A97V mutations showed negative correlation. In addition, among the identified clades, none showed a significant correlation to CFR. The G, GR, GV, S clades showed a significant positive correlation to deaths per million. The GR and S clades showed a positive correlation to number of cases per million. The clades having the highest frequencies among continents were G, followed by GH and GR. Further confirmation and biological significance studies are needed to verify association of identified mutations and clades with a CFR, number of deaths per million and number of cases per million

6. Limitations

Case fatality ratios, number of deaths per million and number of cases per million may be affected by several factors including age distribution differences, blood groups, gender, virus genomic types and ethnic backgrounds; many of these factors still require confirmation.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12071061/s1>, Table S1: Clades frequencies of SARS-CoV-2 genome; Table S2: Correlation analysis between SARS-CoV-2 mutations and Case fatality rates, cases per million, and deaths per million among the identified mutations in the all world continents; Table S3: Dendrogram showing clustering of SARS-CoV-2 isolates and heatmap of the frequency of mutations of the SARS-CoV-2 genome by continent.

Author Contributions: Conceptualization, W.J.A.-A., H.J.A.-A., R.N. and S.S.; methodology, S.E.A.T., R.A., O.A.b., Y.G., M.S.A., A.S., A.M.A., W.J.A.-A., A.I., S.M.A. and B.J.A.H.; validation, N.R.H., W.J.A.-A., S.S. and R.N.; formal analysis, M.W.A., S.S., R.N., A.M.A., W.J.A.-A., S.E.A.T., R.A. and H.J.A.-A.; investigation, S.M.A., S.E.A.T., F.A., R.A., O.A.b., Y.G., S.S., W.J.A.-A. and B.J.A.H.; data curation, S.M.A., M.S.A., S.S., N.R.H., W.J.A.-A. and H.J.A.-A.; writing—original draft preparation, A.S., W.J.A.-A., B.J.A.H., F.A. and H.J.A.-A.; writing—review and editing, S.M.A., A.M.A., R.N., S.M.A., Y.G. and O.A.b. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Peoples' Friendship University of Russia (RUDN University) Strategic Academic Leadership Program.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the American University of Madaba (protocol code: H20010, 25 November 2020).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or Supplementary Material.

Acknowledgments: The authors would like to thank Embry-Riddle Aeronautical University in Prescott-Arizona for covering part of the publication fees.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269. [CrossRef]
2. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [CrossRef]
3. Toyoshima, Y.; Nemoto, K.; Matsumoto, S.; Nakamura, Y.; Kiyotani, K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J. Hum. Genet.* **2020**, *65*, 1075–1082. [CrossRef]
4. Taylor, S.; Landry, C.A.; Paluszczek, M.M.; Fergus, T.A.; McKay, D.; Asmundson, G.J. COVID stress syndrome: Concept, structure, and correlates. *Depress. Anxiety* **2020**, *37*, 706–714. [CrossRef]
5. Lim, Y.X.; Ng, Y.L.; Tam, J.P.; Liu, D.X. Human coronaviruses: A review of virus–host interactions. *Diseases* **2016**, *4*, 26. [CrossRef]
6. Chatterjee, S. Understanding the nature of variations in structural sequences coding for coronavirus spike, envelope, membrane and nucleocapsid proteins of sars-cov-2. In *Envelope, Membrane and Nucleocapsid Proteins of SARS-CoV-2 (28 March 2020)*; 2020. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3562504 (accessed on 28 March 2020).
7. Satarker, S.; Nampoothiri, M. Structural Proteins in Severe Acute Respiratory Syndrome Coronavirus-2. *Arch. Med Res.* **2020**, *51*, 482–491. [CrossRef]
8. Chen, T.; Wu, D.; Chen, H.; Yan, W.; Yang, D.; Chen, G.; Ma, K.; Xu, D.; Yu, H.; Wang, H. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: Retrospective study. *BMJ* **2020**, *368*, M1091. [CrossRef]
9. Subissi, L.; Posthuma, C.C.; Collet, A.; Zevenhoven-Dobbe, J.C.; Gorbalenya, A.E.; Decroly, E.; Snijder, E.J.; Canard, B.; Imbert, I. One severe acute respiratory syndrome coronavirus protein complex integratio processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E3900–E3909. [CrossRef]
10. Van Dorp, L.; Acman, M.; Richard, D.; Shaw, L.P.; Ford, C.E.; Ormond, L.; Owen, C.J.; Pang, J.; Tan, C.C.; Boshier, F.A. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Evol.* **2020**, *83*, 104351. [CrossRef]
11. Yang, W.; Kandula, S.; Huynh, M.; Greene, S.K.; Van Wye, G.; Li, W.; Chan, H.T.; McGibbon, E.; Yeung, A.; Olson, D. Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: A model-based analysis. *Lancet Infect. Dis.* **2021**, *21*, 203–212. [CrossRef]
12. Al-Awaida, W.J.; Al-Ameer, H.J.; Abousenna, M.S.; Nimer, R.; Al-Asafteh, H.O.; Evgenievich, S.A. SARS-Cov-2 Genomic Variations Among Isolates from Jordanian Patients. *Int. J. Pharm. Res.* **2020**, *12*, 3917–3925.
13. Da Silva Filipe, A.; Shepherd, J.G.; Williams, T.; Hughes, J.; Aranday-Cortes, E.; Asamaphan, P.; Ashraf, S.; Balcazar, C.; Bruner, K.; Campbell, A. Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland. *Nat. Microbiol.* **2021**, *6*, 112–122. [CrossRef]
14. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **2018**, *34*, 4121–4123. [CrossRef]
15. Wu, C.; Chen, X.; Cai, Y.; Zhou, X.; Xu, S.; Huang, H.; Zhang, L.; Zhou, X.; Du, C.; Zhang, Y. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern. Med.* **2020**, *180*, 934–943. [CrossRef]
16. Xu, Z.; Shi, L.; Wang, Y.; Zhang, J.; Huang, L.; Zhang, C.; Liu, S.; Zhao, P.; Liu, H.; Zhu, L. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir. Med.* **2020**, *8*, 420–422. [CrossRef]
17. Sudharsanan, N.; Didzun, O.; Bärnighausen, T.; Geldsetzer, P. The Contribution of the Age Distribution of Cases to COVID-19 Case Fatality Across Countries: A Nine-Country Demographic Study. *Ann. Intern. Med.* **2020**, *173*, 714–720. [CrossRef]
18. Boehmer, T.K.; DeVies, J.; Caruso, E.; van Santen, K.L.; Tang, S.; Black, C.L.; Hartnett, K.P.; Kite-Powell, A.; Dietz, S.; Lozier, M. Changing age distribution of the COVID-19 pandemic—United States, May–August 2020. *Morb. Mortal. Wkly. Rep.* **2020**, *69*, 1404. [CrossRef]
19. Göker, H.; Karakulak, E.A.; Demiroğlu, H.; Ceylan, Ç.M.A.; Büyükaşık, Y.; Inkaya, A.Ç.; Aksu, S.; Sayinalp, N.; Haznedaroğlu, I.C.; Uzun, Ö. The effects of blood group types on the risk of COVID-19 infection and its clinical outcome. *Turk. J. Med. Sci.* **2020**, *50*, 679–683. [CrossRef]
20. Alon, T.M.; Doepke, M.; Olmstead-Rumsey, J.; Tertilt, M. *The Impact of COVID-19 on Gender Equality*; National Bureau of Economic Research: Cambridge, MA, USA, 2020.
21. Jin, J.-M.; Bai, P.; He, W.; Wu, F.; Liu, X.-F.; Han, D.-M.; Liu, S.; Yang, J.-K. Gender differences in patients with COVID-19: Focus on severity and mortality. *Front. Public Health* **2020**, *8*, 152. [CrossRef]
22. Pachetti, M.; Marini, B.; Giudici, F.; Benedetti, F.; Angeletti, S.; Ciccozzi, M.; Masciovecchio, C.; Ippodrino, R.; Zella, D. Impact of lockdown on Covid-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America. *J. Transl. Med.* **2020**, *18*, 1–7. [CrossRef]

23. Dreher, N.; Spiera, Z.; McAuley, F.M.; Kuohn, L.; Durbin, J.R.; Marayati, N.F.; Ali, M.; Li, A.Y.; Hannah, T.C.; Gometz, A. Impact of policy interventions and social distancing on SARS-CoV-2 transmission in the United States. *medRxiv* 2020. [[CrossRef](#)]
24. Di Maria, E.; Latini, A.; Borgiani, P.; Novelli, G. Genetic variants of the human host influencing the coronavirus-associated phenotypes (SARS, MERS and COVID-19): Rapid systematic review and field synopsis. *Hum. Genom.* **2020**, *14*, 1–19. [[CrossRef](#)] [[PubMed](#)]
25. Elhabyan, A.; El Yaacoub, S.; Sanad, E.; Mohamed, A.; Elhabyan, A.; Dinu, V. The role of host genetics in susceptibility to severe viral infections in humans and insights into host genetics of severe COVID-19: A systematic review. *Virus Res.* **2020**, *289*, 198163. [[CrossRef](#)]
26. Domingo, E. Mechanisms of viral emergence. *Vet. Res.* **2010**, *41*, 38. [[CrossRef](#)]
27. Rahimi, A.; Mirzazadeh, A.; Tavakolpour, S. Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics* **2020**, *113*, 1221–1232. [[CrossRef](#)] [[PubMed](#)]
28. Holmes, E.C.; Dudas, G.; Rambaut, A.; Andersen, K.G. The evolution of Ebola virus: Insights from the 2013–2016 epidemic. *Nature* **2016**, *538*, 193–200. [[CrossRef](#)]
29. Domingo-Calap, P.; Schubert, B.; Joly, M.; Solis, M.; Untrau, M.; Carapito, R.; Georgel, P.; Caillard, S.; Fafi-Kremer, S.; Paul, N. An unusually high substitution rate in transplant-associated BK polyomavirus in vivo is further concentrated in HLA-C-bound viral peptides. *PLoS Pathog.* **2018**, *14*, e1007368. [[CrossRef](#)]
30. Pachetti, M.; Marini, B.; Benedetti, F.; Giudici, F.; Mauro, E.; Storici, P.; Masciovecchio, C.; Angeletti, S.; Ciccozzi, M.; Gallo, R.C. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.* **2020**, *18*, 1–9. [[CrossRef](#)] [[PubMed](#)]
31. Chen, Z.-W.; Li, Z.; Li, H.; Ren, H.; Hu, P. Global genetic diversity patterns and transmissions of SARS-CoV-2. *medRxiv* 2020. [[CrossRef](#)]
32. Zhao, Z.; Li, H.; Wu, X.; Zhong, Y.; Zhang, K.; Zhang, Y.-P.; Boerwinkle, E.; Fu, Y.-X. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol. Biol.* **2004**, *4*, 21. [[CrossRef](#)]
33. Chan, J.F.-W.; Kok, K.-H.; Zhu, Z.; Chu, H.; To, K.K.-W.; Yuan, S.; Yuen, K.-Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* **2020**, *9*, 221–236. [[CrossRef](#)]
34. Zhu, N.; Wang, W.; Liu, Z.; Liang, C.; Wang, W.; Ye, F.; Huang, B.; Zhao, L.; Wang, H.; Zhou, W. Morphogenesis and cytopathic effect of SARS-CoV-2 infection in human airway epithelial cells. *Nat. Commun.* **2020**, *11*, 1–8. [[CrossRef](#)] [[PubMed](#)]
35. Laha, S.; Chakraborty, J.; Das, S.; Manna, S.K.; Biswas, S.; Chatterjee, R. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect. Genet. Evol.* **2020**, *85*, 104445. [[CrossRef](#)]
36. Laamarti, M.; Alouane, T.; Kartti, S.; Chemaou-Elfihri, M.; Hakmi, M.; Essabbar, A.; Laamart, M.; Hlali, H.; Allam, L.; El Hafidi, N. Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geodistribution and a rich genetic variations of hotspots mutations. *bioRxiv* **2020**, *15*, e0240345.
37. Wen, F.; Yu, H.; Guo, J.; Li, Y.; Luo, K.; Huang, S. Identification of the hyper-variable genomic hotspot for the novel coronavirus SARS-CoV-2. *J. Infect.* **2020**, *80*, 671–693. [[CrossRef](#)] [[PubMed](#)]
38. Soratto, T.A.T.; Darban, H.; Bjerkner, A.; Coorens, M.; Albert, J.; Allander, T.; Andersson, B. Four SARS-CoV-2 genome sequences from late April in Stockholm, Sweden, reveal a rare mutation in the spike protein. *Microbiol. Resour. Announc.* **2020**, *9*, e00934-20. [[CrossRef](#)] [[PubMed](#)]
39. Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B. Tracking changes in SARS-CoV-2 Spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **2020**, *182*, 812–827. [[CrossRef](#)]
40. Grubaugh, N.D.; Hanage, W.P.; Rasmussen, A.L. Making sense of mutation: What D614G means for the COVID-19 pandemic remains unclear. *Cell* **2020**, *182*, 794–795. [[CrossRef](#)] [[PubMed](#)]
41. Hou, Y.J.; Chiba, S.; Halfmann, P.; Ehre, C.; Kuroda, M.; Dinno, K.H.; Leist, S.R.; Schäfer, A.; Nakajima, N.; Takahashi, K. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **2020**, *370*, 1464–1468. [[CrossRef](#)]
42. Gordon, D.E.; Jang, G.M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K.M.; O’Meara, M.J.; Rezelj, V.V.; Guo, J.Z.; Swaney, D.L. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583*, 459–468. [[CrossRef](#)]
43. Rosas-Lemus, M.; Minasov, G.; Shuvalova, L.; Inniss, N.L.; Kiryukhina, O.; Wiersum, G.; Kim, Y.; Jedrzejczak, R.; Maltseva, N.I.; Endres, M. The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv* **2020**. [[CrossRef](#)]
44. Angeletti, S.; Benvenuto, D.; Bianchi, M.; Giovanetti, M.; Pascarella, S.; Ciccozzi, M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J. Med. Virol.* **2020**, *92*, 584–588. [[CrossRef](#)]
45. Madu, I.G.; Belouzard, S.; Whittaker, G.R. SARS-coronavirus spike S2 domain flanked by cysteine residues C822 and C833 is important for activation of membrane fusion. *Virology* **2009**, *393*, 265–271. [[CrossRef](#)] [[PubMed](#)]
46. Zheng, Q.; Deng, Y.; Liu, J.; Van Der Hoek, L.; Berkhout, B.; Lu, M. Core structure of S2 from the human coronavirus NL63 spike glycoprotein. *Biochemistry* **2006**, *45*, 15205–15215. [[CrossRef](#)]
47. Kang, S.; Yang, M.; Hong, Z.; Zhang, L.; Huang, Z.; Chen, X.; He, S.; Zhou, Z.; Zhou, Z.; Chen, Q. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* **2020**, *10*, 1228–1238. [[CrossRef](#)]

48. Sheikh, A.; Al-Taher, A.; Al-Nazawi, M.; Al-Mubarak, A.I.; Kandeel, M. Analysis of preferred codon usage in the coronavirus N genes and their implications for genome evolution and vaccine design. *J. Virol. Methods* **2020**, *277*, 113806. [[CrossRef](#)]
49. Chang, C.-K.; Hou, M.-H.; Chang, C.-F.; Hsiao, C.-D.; Huang, T.-H. The SARS coronavirus nucleocapsid protein—forms and functions. *Antivir. Res.* **2014**, *103*, 39–50. [[CrossRef](#)]
50. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Glob. Chall.* **2017**, *1*, 33–46. [[CrossRef](#)] [[PubMed](#)]
51. Alm, E.; Broberg, E.K.; Connor, T.; Hodcroft, E.B.; Komissarov, A.B.; Maurer-Stroh, S.; Melidou, A.; Neher, R.A.; O’Toole, Á.; Pereyaslov, D. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Eurosurveillance* **2020**, *25*, 2001410. [[CrossRef](#)]
52. Raghav, S.; Ghosh, A.; Turuk, J.; Kumar, S.; Jha, A.; Madhulika, S.; Priyadarshini, M.; Biswas, V.K.; Shyamli, P.S.; Singh, B. Analysis of Indian SARS-CoV-2 Genomes Reveals Prevalence of D614G Mutation in Spike Protein Predicting an Increase in Interaction with TMPRSS2 and Virus Infectivity. *Front. Microbiol.* **2020**, *11*, 594928. [[CrossRef](#)] [[PubMed](#)]
53. Brufsky, A. Distinct viral clades of SARS-CoV-2: Implications for modeling of viral spread. *J. Med. Virol.* **2020**, *92*, 1386–1390. [[CrossRef](#)] [[PubMed](#)]
54. Walls, A.C.; Park, Y.-J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veessler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**, *181*, 281–292.e286. [[CrossRef](#)] [[PubMed](#)]