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.

Environmental Research 204 (2022) 112092

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Periodically aperiodic pattern of SARS-CoV-2 mutations underpins the uncertainty of its origin and evolution

Sk Sarif Hassan [a,**], Pallab Basu [b], Elrashdy M. Redwan [c,d], Kenneth Lundstrom [e,*], Pabitra Pal Choudhury [f], Ángel Serrano-Aroca [g], Gajendra Kumar Azad [h], Alaa A.A. Aljabali [i], Giorgio Palu [j], Tarek Mohamed Abd El-Aziz [k,l], Debmalya Barh [m,n], Bruce D. Uhal [o], Parise Adadi [p], Kazuo Takayama [q], Nicolas G. Bazan [r], Murtaza M. Tambuwala [s], Amos Lal [t], Gaurav Chauhan [u], Wagner Baetas-da-Cruz [v], Samendra P. Sherchan [w], Vladimir N. Uversky [x,y,***]

[a] Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram, Paschim Medinipur, 721140, West Bengal, India
[b] School of Physics, University of the Witwatersrand, Johannesburg, Braamfontein 2000, 721140, South Africa
[c] Biological Science Department, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia
[d] Therapeutic and Protective Proteins Laboratory, Protein Research Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications, New Borg EL-Arab, 21934, Alexandria, Egypt
[e] PanTherapeutics, Rte de Lavaux 49, CH1095, Lutry, Switzerland
[f] Indian Statistical Institute, Applied Statistics Unit, 203 B T Road, Kolkata, 700108, India
[g] Biomaterials & Bioengineering Lab, Centro de Investigación Traslacional San Alberto Magno, Universidad Católica de Valencia, San Vicente Mártir, Valencia 46001, Spain
[h] Department of Zoology, Patna University, Patna, Bihar, India
[i] Department of Pharmaceutics and Pharmaceutical Technology, Yarmouk University, Faculty of Pharmacy, Irbid, 566, Jordan
[j] Department of Molecular Medicine, University of Padova, Via Gabelli 63, 35121, Padova, Italy
[k] Zoology Department, Faculty of Science, Minia University, El-Minia, 61519, Egypt
[l] Department of Cellular and Integrative Physiology, University of Texas Health Science Center at San Antonio, San Antonio, TX, 78229-3900, USA
[m] Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, WB, India
[n] Departamento de Geńetica, Ecologia e Evolucao, Instituto de Cĩencias Biólogicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[o] Department of Physiology, Michigan State University, East Lansing, MI, 48824, USA
[p] Department of Food Science, University of Otago, Dunedin, 9054, New Zealand
[q] Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, 6068507, Japan
[r] Neuroscience Center of Excellence, School of Medicine, LSU Health New Orleans, New Orleans, LA, 70112, USA
[s] School of Pharmacy and Pharmaceutical Science, Ulster University, Coleraine, BT52 1SA, Northern Ireland, UK
[t] Department of Medicine, Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, MN, USA
[u] School of Engineering and Sciences, Tecnologico de Monterrey, Av. Eugenio Garza Sada 2501 Sur, 64849, Monterrey, Nuevo Léon, Mexico
[v] Translational Laboratory in Molecular Physiology, Centre for Experimental Surgery, College of Medicine, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
[w] Department of Environmental Health Sciences, Tulane University, New Orleans, LA, 70112, USA
[x] Department of Molecular Medicine and USF Health Byrd Alzheimer's Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA
[y] Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Institutskiy pereulok, 9, Dolgoprudny, 141700, Russia

* Corresponding author.
** Corresponding author.
*** Corresponding author. Department of Molecular Medicine and USF Health Byrd Alzheimer's Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA.

E-mail addresses: sksarifhassan@pinglacollege.ac.in (S.S. Hassan), pallabbasu@gmail.com (P. Basu), lradwan@kau.edu.sa (E.M. Redwan), lundstromkenneth@gmail.com (K. Lundstrom), pabitrapalchoudhury@gmail.com (P.P. Choudhury), angel.serrano@ucv.es (Á. Serrano-Aroca), gkazad@patnauniversity.ac.in (G.K. Azad), alaaj@yu.edu.jo (A.A.A. Aljabali), giorgio.palu@unipd.it (G. Palu), mohamedt1@uthscsa.edu (T.M. Abd El-Aziz), dr.barh@gmail.com (D. Barh), bduhal@gmail.com (B.D. Uhal), pariseadadi@gmail.com (P. Adadi), kazuo.takayama@cira.kyoto-u.ac.jp (K. Takayama), nbazan@lsuhsc.edu (N.G. Bazan), m.tambuwala@ulster.ac.uk (M.M. Tambuwala), manavamos@gmail.com (A. Lal), gchauhan@tec.mx (G. Chauhan), wagner.baetas@gmail.com (W. Baetas-da-Cruz), sshercha@tulane.edu (S.P. Sherchan), vuversky@usf.edu (V.N. Uversky).

https://doi.org/10.1016/j.envres.2021.112092
Received 22 June 2021; Received in revised form 17 September 2021; Accepted 18 September 2021
Available online 22 September 2021
0013-9351/© 2021 Elsevier Inc. All rights reserved.

ARTICLE INFO

Keywords:
SARS-CoV-2
Invariant residues
Mutations
Relative frequency
Aperiodically periodic

ABSTRACT

Various lineages of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) have contributed to prolongation of the Coronavirus Disease 2019 (COVID-19) pandemic. Several non-synonymous mutations in SARS-CoV-2 proteins have generated multiple SARS-CoV-2 variants. In our previous report, we have shown that an evenly uneven distribution of unique protein variants of SARS-CoV-2 is geo-location or demography-specific. However, the correlation between the demographic transmutability of the SARS-CoV-2 infection and mutations in various proteins remains unknown due to hidden symmetry/asymmetry in the occurrence of mutations. This study tracked how these mutations are emerging in SARS-CoV-2 proteins in six model countries and globally. In a geo-location, considering the mutations having a frequency of detection of at least 500 in each SARS-CoV-2 protein, we studied the country-wise percentage of invariant residues. Our data revealed that since October 2020, highly frequent mutations in SARS-CoV-2 have been observed mostly in the Open Reading Frame (ORF) 7b and ORF8, worldwide. No such highly frequent mutations in any of the SARS-CoV-2 proteins were found in the UK, India, and Brazil, which does not correlate with the degree of transmissibility of the virus in India and Brazil. However, we have found a signature that SARS-CoV-2 proteins were evolving at a higher rate, and considering global data, mutations are detected in the majority of the available amino acid locations. Fractal analysis of each protein's normalized factor time series showed a periodically aperiodic emergence of dominant variants for SARS-CoV-2 protein mutations across different countries. It was noticed that certain high-frequency variants have emerged in the last couple of months, and thus the emerging SARS-CoV-2 strains are expected to contain prevalent mutations in the ORF3a, membrane, and ORF8 proteins. In contrast to other beta-coronaviruses, SARS-CoV-2 variants have rapidly emerged based on demographically dependent mutations. Characterization of the periodically aperiodic nature of the demographic spread of SARS-CoV-2 variants in various countries can contribute to the identification of the origin of SARS-CoV-2.

1. Introduction

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is the causative agent of the Coronavirus Disease (COVID-19) (Pedersen and Ho, 2020; Setti et al., 2020; Domingo et al., 2020). SARS-CoV-2 has spread rapidly and has evolved prolonging pandemic and precarious clinical entity (Health, 2020; Tapper and Asrani, 2020). Since the beginning of the pandemic, SARS-CoV-2 has increasingly accumulated various mutations leading to patterns of genomic diversity (van Dorp et al., 2020; Zhou et al., 2020). The wide SARS-CoV-2 variations were scattered across the various geo-locations, and it can underlie geographically specific etiological effects (Mercatelli and Giorgi, 2020). It was expected that these mutations could be of use to monitor the spread of the virus, and to identify sites putatively under selection as SARS-CoV-2 potentially adapts to its new human host. SARS-CoV-2 may be evolving towards higher transmissibility as it may not yet fully adapt to its human host. The most plausible mutations under putative natural selection are those which have emerged repeatedly and independently (van Dorp et al., 2020). It was reported that 198 sites in the SARS-CoV-2 genome appear to have already undergone recurrent, independent mutations (van Dorp et al., 2020). Various SARS-CoV-2 missense mutations are the key evolving factors affecting the infectivity, and virulence, and pathogenicity of the virus (Mercatelli and Giorgi, 2020). Several SARS-CoV-2 variants have significantly strengthened the infectivity (Chen et al., 2020). It was previously reported that the rate of SARS-CoV-2 mutations are relatively low compared to other RNA viruses, such as influenza virus (Kupferschmidt, 2020; Khan et al., 2020; Gómez-Carballa et al., 2020). The low SARS-CoV-2 mutation rate might relate to its proofreading ability, which is a unique embedded function of SARS-CoV-2 (Romano et al., 2020; Ogando et al., 2020). Thus far, although several mutations have been detected, SARS-CoV-2 seem not to be drifting antigenically (Yuan et al., 2021; Williams and Burgers, 2021). However, the mechanism of SARS-CoV-2 evolution or developing gain of function variations have remained unclear (Wang et al., 2021). A non-uniform mutation pattern in the viral proteins was recently reported, which has further accelerated the discussion on the question of the origin of SARS-CoV-2 (Hassan et al., 2021b).

The rapidly evolving data on mutations and various strains of SARS-CoV-2 makes it vulnerable to firmly assert whether SARS-CoV-2 results from a zoonotic emergence or from an accidental escape from a laboratory (Sallard et al., 2021; Pipes et al., 2021; Nadeau et al., 2021; Seyran et al., 2021). This issue of origin needs to be resolved because it has important consequences on the risk/benefit balance of human interactions with ecosystems, on intensive breeding of wild and domestic animals, on some laboratory practices and on scientific policy and bio-safety regulations (Sallard et al., 2021). Despite these recent investigations, several issues related to the evolutionary patterns and origin of the COVID-19 pandemic remain to be fully characterized (Liu et al., 2020, 2019; Domingo, 2021a,b; Grassberger, 1993). No direct correlation was observed in the mutation pattern of SARS-CoV-2 from the infection rate in the first and second waves of COVID-19 (Ko et al., 2021; Lv et al., 2020; Kumar et al., 2020).

In this study, the potential embedded mutation pattern of the spike (S), envelope (E), membrane (M) (Domingo, 2021b), nucleocapsid (N), ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 proteins of SARS-CoV-2 are analyzed in six model countries viz. USA, UK, Brazil, Germany, India, South Africa (SA), and globally.

2. Data specifications and methods

2.1. Data

The majority of publicly available SARS-CoV-2 genomic sequences were sourced from GISAID, NCBI,and CNGB. The SARS-CoV-2 sequence data was taken from the GISAID database (Shu and McCauley, 2017). Mutation data with their respective details were collected from the CoVal database. In this study, we focused on the S, M, E, N, ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 proteins of SARS-CoV-2. We also considered mutations within the protein of our interest from all geo-locations available in the CoVal database. In particular, a set of six model countries, the USA, UK, South Africa, India, Germany, and Brazil was selected.

2.2. Methods

2.2.1. Mutation search using the CoVal database

Single mutation details were retrieved from the CoVal database by searching by the name of the model country and the SARS-CoV-2 protein of interest. For example, details of all single mutations in the SARS-CoV-2 S protein from the UK were retrieved, of which a snapshot is presented in Figure 1.

Likewise, for other model countries and SARS-CoV-2 proteins, details

of single mutations were retrieved. Prior to proceeding into the result section, some definitions are recalled and redefined for easy reading.

3. Results

Four different classes were defined on the date of the first detection and frequency of mutations in the SARS-CoV-2 proteins of our interest. Class-I contains all mutations detected in the proteins of SARS-CoV-2 across the world. Class-II contains only those mutations with a frequency more than or equal to 500 (a reasonably good frequency of a mutation detected in a geo-location) in SARS-CoV-2 in affected patients worldwide. Mutations that were detected after October 2020 belong to Class-III. Mutations with a frequency larger than 500 since October 2020 are members of Class-IV.

3.1. Invariant residues of SARS-CoV-2 proteins

Amino acid residues where no mutations were detected are termed "invariant residues". From the CoVal database, we first found distinct residue positions of mutations, and the total number of invariant residues (r) of each type. Furthermore, the percentage of invariant residues in each protein of length (l) was determined using the formula $^{l-r} \times 100$.

Class-I. The percentages of invariant residues (Class-I) in SARS-CoV-2 proteins in various countries are listed in Table 1.

Considering all mutations with amino acid changes in all available geo-locations, it was observed that, except for the SARS-CoV-2 structural proteins, the ORF proteins (ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10) possessed mutations at every residue position of the respective protein. On the other hand, it was found that an increasing order of the percentage of invariant residues in the structural proteins of SARS-CoV-2 turned out to be E (6.67) < N (10.74) < S (14.69) < M (24.32). In other words, the highest and lowest number of mutations were detected in the E and M proteins, respectively. Across six countries, the highest and lowest number of invariant residues were found in the M and ORF3a proteins, respectively. It was noted that the highest frequency of invariant residues in the E among all proteins was observed in Germany. Among all proteins, the highest number of mutations was detected in the ORF3a proteins in Germany, India, Brazil, and South Africa, whereas in the USA and UK, ORF7b possessed the highest number of mutations. Notably, it was observed that each protein in SARS-CoV-2 possessed an almost similar number of mutations in the USA and UK. Among six countries, the least amount of mutations across all proteins was found in COVID-19 patients from South Africa, whereas in the USA, the highest number of mutations in SARS-CoV-2 were detected.

An increasing-order (decreasing-order) of the six geo-locations based on the invariant residues (mutations) across all proteins was found in the USA < UK < Germany < India < Brazil < SA (reverse order).

Class-II. The percentages of invariant residues (Class-II) over the SARS-CoV-2 proteins in various countries are listed in Table 2.

While we considered only those mutations with a frequency of at least 500, it was observed that the increasing order of the percentage of invariant residues in the SARS-CoV-2 structural proteins was N (80.19) < S (89.94)< E (92) < M (94.14). Therefore, the highest and lowest frequency of mutations were detected in the N and M proteins, respectively. Across six geo-locations, the ORF3a and M proteins possessed the highest and lowest amount of amino acid changing mutations, respectively. It was further noticed that each SARS-CoV-2 protein possessed almost the same number of mutations in the USA and UK. Among six countries, the least number of mutations across all proteins is found in COVID-19 patients from the UK, whereas the highest number of mutations were detected in India.

A decreasing order of the total number of mutations in proteins at the six geo-locations was the UK > USA > Germany > SA > Brazil > India (from highest to lowest). In other words, the highest number of different

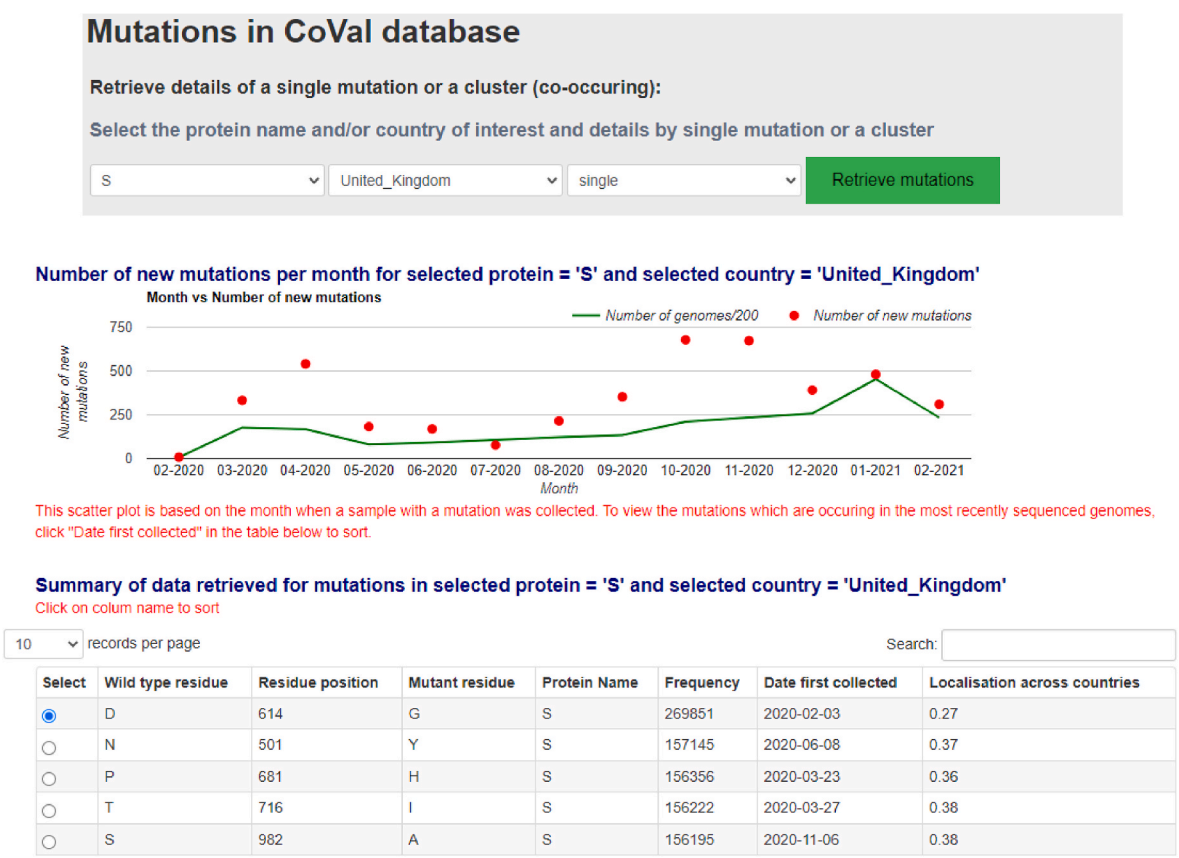

Fig. 1. Details of a single mutation retrieval from the CoVal database.

Table 1
Percentage of invariant residues (considering all mutations so far accounted) in the SARS-CoV-2 proteins in various countries.

Countries	S	E	M	N	ORF10	ORF3a	ORF6	ORF7a	ORF7b	ORF8
Across all countries	14.69	6.67	24.32	10.74	0	1.1	0	0	0	0
USA	32.52	29.34	37.39	19.33	5.26	5.1	6.56	2.48	2.32	2.48
UK	33.54	29.34	42.8	19.57	7.9	7.63	6.55	1.65	0	4.13
Germany	61.35	72	69.82	46.06	44.74	24.37	39.35	71.07	34.99	27.27
India	80.83	80	88.29	69.21	57.89	55.64	77.05	66.11	55.81	60.33
Brazil	87.51	84	92.8	80.67	78.95	71.28	83.6	73.55	76.74	73.55
SA	90.49	90.67	94.14	82.82	84.21	77.82	91.8	90.08	83.72	82.64

Table 2
Percentage of invariant residues (considering all mutations having a frequency >500) in the SARS-CoV-2 proteins in various countries.

Countries	S	E	M	N	ORF10	ORF3a	ORF6	ORF7a	ORF7b	ORF8
Across all countries	89.94	92	94.14	80.19	78.95	72	91.8	81	83.72	72.72
USA	96.54	97.34	99.55	92.6	94.74	92.36	100	98.35	97.67	89.26
UK	97.1	100	99.1	93.8	94.74	89.1	98.36	96.7	93.02	93.39
Germany	98.74	100	100	97.61	97.37	97.45	100	100	100	98.35
India	99.92	100	100	99.28	100	99.63	100	100	100	100
Brazil	99.76	100	100	99.04	100	100	98.36	100	100	100
SA	99.53	98.67	100	99.28	100	99.27	100	100	100	100

mutations across all these proteins were detected in the UK, and the lowest number of mutations was observed in India. In India, Brazil, and SA, no mutation with a frequency of at least 500 (originating in these countries) was detected in any SARS-CoV-2 protein, except one or two reported most frequent deleterious mutations D614G in the S protein and Q57H in the ORF3a protein. Note that in Germany, the S, N, ORF10, ORF3a, and ORF8 proteins possessed a couple of mutations, whereas the E, M, ORF6, ORF7a, and ORF7b proteins did not have any mutations originating from Germany, with a frequency of more than 500.

Class-III. The percentages of invariant residues (Class-III) in the SARS-CoV-2 proteins in various countries are presented in Table 3.

It was observed that the SARS-CoV-2 ORF3a protein was no longer a hotspot for dominant mutations. From October 2020 until June 2021, the M and ORF8 proteins owned the lowest (66.22%) and the highest (99.174%) number of mutations, respectively, across the world. Notably, in the past, the ORF3a protein possessed the highest number of mutations (Bianchi et al., 2021; Hassan et al., 2020). Currently, it seems that ORF3a mutations in the USA, UK, and elsewhere are relatively rare (less than 10%). Since October 2020, the highest number of mutations have been detected in the ORF3a protein in SA. In the UK, Germany, India, and SA, the E protein had the lowest frequency of mutations, whereas the M protein possessed several mutations amounting to the highest frequency among others in the USA and Brazil. The highest percentage of mutations were detected in the ORF8 protein, since October 2020 in Germany, India, and Brazil. The highest frequency of mutations in the USA and UK were observed in the ORF7a and ORF7b proteins, respectively.

An increasing order of the six geo-locations based on the variability of mutations in SARS-CoV-2 proteins (from October 2020 until June 2021) was India < SA < Brazil < Germany < USA < UK.

Class-IV. The percentages of invariant residues (Class-IV) in SARS-CoV-2 proteins in various countries are listed in Table 4.

Considering all mutations having a frequency of more than 500, from October 2020 until June 2021, it was observed that the ORF7b and ORF8 proteins owned the highest frequency of various mutations, whereas the M protein contained the lowest number of mutations. In fact, all four structural proteins S, E, M, and N contain more than 94% of invariant residues. No mutations that had a frequency of more than 500 since October 2020 in the UK, India, and Brazil were found in any of the aforementioned SARS-CoV-2 proteins. In South Africa, the proteins S, E, N, ORF3a, and ORF10 owned a couple of mutations with a frequency of more than 500 detected since October 2020. The S, N, and ORF10 proteins possessed a couple of mutations of this kind in Germany, whereas only in the S and E proteins, a few mutations of this type were noticed.

3.2. *Hidden self-organized self-similarity in frequency of SARS-CoV-2 mutations*

The degree of fractality, fractal dimension (FD), of the normalized factor time series (NFTS) for each SARS-CoV-2 protein in a given model country was quantified (Table 5). The FD revealed any hidden self-organized self-similar patterns (if any) in the frequency of emerging mutations in each protein (Pilgrim and Taylor, 2018; Sánchez-Granero et al., 2012). Here the NFTS is defined as follows:

Normalized factor time series: For each mutation (m) in a given SARS-CoV-2 protein, a normalized factor is determined based on the presence of the mutation in various geo-locations (countries). Normalized factor NF_m is defined as

$$NF_m = \frac{NG_m}{TG_m}$$

Table 3
Percentage of invariant residues (considering all mutations since October 2020) in the SARS-CoV-2 proteins in various countries.

Countries	S	E	M	N	ORF10	ORF3a	ORF6	ORF7a	ORF7b	ORF8
Across all countries	24.51	22.67	33.78	15.03	2.63	4	1.64	0.826	4.65	0.826
USA	56.72	56	64.41	51.55	18.42	33.1	32.79	26.45	13.95	28.92
UK	53.18	61.33	57.2	49.4	55.26	28.73	16.39	9.92	23.25	23.14
Germany	68.66	82.67	77.48	67.3	60.53	40.73	47.54	39.67	41.86	37.2
India	97.49	100	97.75	97.14	92.1	96	98.36	93.39	95.35	91.73
Brazil	94.58	94.67	97.75	94.03	97.37	90.18	91.8	90.9	95.35	88.43
SA	97.25	100	98.65	94.99	97.37	92.37	98.36	97.52	93.02	95.87

Table 4
Percentage of invariant residues (considering all mutations since October 2020 having a frequency of >500) in the SARS-CoV-2 proteins in various countries.

Countries	S	E	M	N	ORF10	ORF3a	ORF6	ORF7a	ORF7b	ORF8
Across all countries	95.68	96	98.2	94.51	92.1	90.91	96.72	95.04	88.37	88.43
USA	99.76	98.67	100	100	100	100	100	100	100	100
UK	100	100	100	100	100	100	100	100	100	100
Germany	99.6	100	100	98.8	97.37	100	100	100	100	100
India	100	100	100	100	100	100	100	100	100	100
Brazil	100	100	100	100	100	100	100	100	100	100
SA	99.53	98.67	100	99.76	97.37	99.27	100	100	100	100

Table 5
Fractal dimension of the normalized factor time series.

Countries	S	E	M	N	ORF3a	ORF6	ORF7a	ORF7b	ORF8	ORF10
All	*1.9336*	*1.7811*	*1.9167*	*1.9351*	*1.9272*	*1.916*	*1.9158*	*1.9396*	*1.9216*	*1.6707*
UK	1.9496	1.603	1.9131	1.9352	1.9146	1.899	1.923	1.8826	1.9352	1.6439
India	1.7112	1.1699	1.2521	1.7539	1.7603	1.0925	1.4771	1.179	1.7549	1.179
USA	1.9484	1.8502	1.9333	1.9308	1.9271	1.9302	1.9232	1.9376	1.9156	1.9771
SA	1.6752	0.6491	1	1.585	1.585	0.5144	1.4594	0.6491	1.179	0.5144
Germany	1.8155	1.179	1.5319	1.8507	1.7984	1.7618	1.7379	1.3315	1.7297	1.1377
Brazil	1.8119	1.585	1.0925	1.7655	1.6705	1.585	1.2849	1.3219	1.5502	0.6491

Here, NG_m and TG_m denote the number of genomes with this specific mutation (m) in this country, and the total number of genomes with this mutation (m) worldwide, respectively. It varies from 0 to 1. Here the normalized factor 0 denotes the spreading of the mutations uniformly across various geo-locations, whereas 1 denotes the detection of the mutation in a single geo-location (Gelly et al., 2011). A series of detected mutations in a given protein lead to anNFTS.

From Table 5 it was concluded that the FD of NFTS for SARS-CoV-2 proteins ranged from 0.5144 to 1.9771. This wide range of the FD depicts the hidden non-uniform distribution of the frequency in various countries. We observed that the lowest FD for each SARS-CoV-2 protein was detected in South Africa, whereas the highest fractality for all proteins were either found in the USA or UK. These model countries were ranked in increasing order based on the FD for each protein (Table 6).

The ranking of countries based on the FD of NFTS for the S and N proteins was identical (Table 6). Likewise, an identical ranking was noticed for the SARS-CoV-2 M and ORF3a proteins. Also, the country-ranking based on the FD of NFTS for the ORF6 and ORF7b proteins were the same. Notably, the fractality of NFTS for the E and ORF7a proteins in India, Germany and Brazil did not follow any strict order.

3.3. The emergence of new SARS-COV-2 variants

To understand the emergence of new strains and which SARS-CoV-2 strain dominates, we looked at the dates of the detection of various mutations in each SARS-CoV-2 protein and their respective frequencies (in logarithmic scale) (Table 7).

A negative correlation signifies that the early detected mutations since the beginning of the COVID-19 pandemic, are more dominant in the present scenario (Lee Rodgers and Nicewander, 1988). We noticed precisely that the pattern in the USA, UK, and Germany was that each correlation coefficient was negative (Mukaka, 2012). The similar pattern was seen in South Africa and Brazil. We detected newer ORF3a mutations that were dominant in South Africa, and some new ORF10 mutations were seen. The trend was that the newer ORF10 mutations were more dominant in Brazil. Mutations in the E and S proteins in Brazil also showed the emergence of new strains. To understand the emergence of newer variants, we plotted the date and frequencies (log) of mutations in the SARS-CoV-2 E, M, ORF3a, and ORF8 proteins (Figure 2). Outlier points (in red) in Figures 2 and 3 showed the emergence of newer mutations in the four SARS-CoV-2 proteins.

The density of outlier points in the ORF3a protein implies the emergence of new variants of SARS-CoV-2 with a significant number of missense mutations. These new SARS-CoV-2 variants would also contain various mutations in the M and ORF8 proteins, whereas it was observed that mutations in the E protein were no longer dominant.

The first twenty dominant (with regards to frequency) mutations worldwide are plotted for all proteins in Figure 3.

It seems that the M protein had more versatile dominant mutants globally (Fig. 3). For the ORF10 protein, only one dominant mutant P10 was identified globally. Two mutants (S24, R52) dominate globally for the ORF8 protein. For the E and S proteins, there was one dominant mutant in each protein and many sub-dominant mutants. The N protein showed an intermediary pattern between the M and the ORF8, ORF10, ORF6, and ORF7a proteins and it also falls into the common pattern of a few dominant mutants and a dozen or so sub-dominant mutants.

Table 6
Increased order (from top to bottom row) of countries based on the fractal dimension (FD) of the normalized factor time series. Identical rankings of countries based on the FD of normalized factor time series (NFTS) are highlighted in red, ochre, and green colors. Fractality of NFTS that did not follow any strict ordering are in black.

S	E	M	N	ORF3a	ORF6	ORF7a	ORF7b	ORF8	ORF10
SA	SA	SA	SA	SA	SA	Brazil	SA	SA	SA
India	India	Brazil	India	Brazil	India	SA	India	Brazil	Brazil
Brazil	Germany	India	Brazil	India	Brazil	India	Brazil	Germany	Germany
Germany	Brazil	Germany	Germany	Germany	Germany	Germany	Germany	India	India
USA	UK	UK	USA	UK	UK	UK	UK	USA	UK
UK	USA	USA	UK	USA	USA	USA	USA	UK	USA

Table 7
Correlation of the date a mutation was first detected and its frequency (log).

Countries	S	E	M	N	ORF10	ORF3a	ORF6	ORF7a	ORF7b	ORF8
All	−0.20	0.01	−0.25	−0.31	−0.21	−0.25	−0.22	−0.14	−0.07	−0.27
UK	−0.46	−0.41	−0.49	−0.47	−0.55	−0.50	−0.43	−0.36	−0.47	−0.47
India	−0.22	−0.13	−0.22	−0.42	−0.24	−0.35	0.29	−0.20	−0.37	−0.47
USA	−0.51	−0.39	−0.47	−0.54	−0.55	−0.53	−0.46	−0.45	−0.50	−0.55
SA	−0.82	−0.99	0.14	−0.39	0.15	−0.68	0.72	0.30	−0.39	−0.38
Germany	−0.34	−0.39	−0.25	−0.33	−0.71	−0.18	−0.32	−0.36	−0.40	−0.47
Brazil	−0.09	−0.06	−0.22	−0.22	0.48	−0.31	−0.49	−0.37	0.36	0.11

Fig. 2. Date and frequencies (log) of mutations in E, M, ORF3a, and ORF8 proteins in the USA.

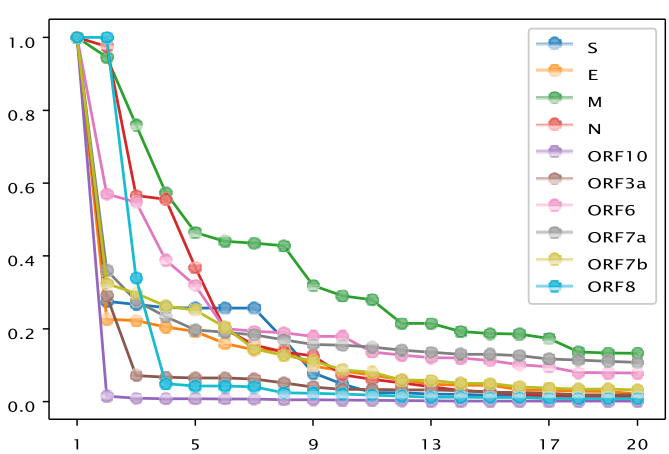

Fig. 3. Relative frequency of the first 20 dominant mutations of S, E, M, N, ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 proteins of SARS-CoV-2.

4. Discussion and concluding remarks

We studied and tracked how mutant variants emerged in the SARS-CoV-2 S, E, M, N, ORF10, ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 proteins from six model countries. While we considered all SARS-CoV-2 mutations since the beginning of the COVID-19 pandemic, it was noticed that the least number of invariant residues in each SARS- CoV-2 protein was observed in South Africa. On the other hand, prevalent missense mutations were detected in the UK and USA. Considering the mutations in each protein having a frequency of more than 500, it was found that the increasing order of the model countries based on the percentage of invariant residues is UK < USA < Germany < SA < Brazil < India. Interestingly, India had the lowest number of protein variants with high frequent mutations (frequency of more than 500). Furthermore, the ample protein variations in the USA and UK were ascribed to the presence of highly frequent mutations in each SARS-CoV-2 protein. Since October 2020, high-frequent mutations in SARS-CoV-2 have been found only in ORF7b and ORF8, worldwide. No such highly frequent mutations in any of the SARS-CoV-2 proteins were found in the UK, India, and Brazil, which does not seem to correlate with the trans-mutability of the virus in India and Brazil. Not every high-frequency mutation necessarily affects the life cycle of the virus.Neither ORF7b nor ORF8 proteins are essential for the virus life cycle as previously

shown for the S and/or ORF3a proteins (Hassan et al., 2021a, b). Of note, the majority of Manaus (Brazil) citizens had been reinfected with the P.1 strain bearing these mutations, which may offer a further capability of immune escape (Faria et al., 2021). The combination of the E484Kand Δ140 mutations led to a complete abolition of neutralizing antibodies. Based on these observations the questionis whether there is any ethnic correlation with SARS-CoV-2 infection? Ethnic/racial disparities in the SARS-CoV-2 infection, incidence, and COVID-19 severity and deaths have been reported (Hollis et al., 2021; McCoy et al., 2020; Vahidy et al., 2020; Shoily et al., 2021; Sze et al., 2020). In the context of correlation between ethnic/racial origin and SARS-CoV-2 infections the interaction between the S protein and the ACE2-solute carrier family 6-member 19 (SLC6A19) dimer was investigated applying a quantitative dynamics cross-correlation matrix (Gupta et al., 2020). Forty-seven potential functional missense variants were identified from genomic databases within ACE2/SLC6A19/TMPRSS2, which justified genomic enrichment studies in COVID-19 patients. Although the variants occurred at ultra-low frequency, two ACE2 non-coding variants (rs4646118 and rs143185769) found in 9% of African Americans may be involved in the regulation of ACE2 expression and may contribute to an enhanced risk of COVID-19. Additionally, studies have demonstrated that ACE2 expression levels relate to the rs2285666 polymorphism significantly affecting SARS-CoV-2 susceptibility in Asian and European populations (Asselta et al., 2020; Srivastava et al., 2020). Moreover, it was discovered that SARS-CoV-2 cell entry depended on polymorphism in the ACE2, TMPRSS2, TMPRSS11A, cathepsin L (CTSL), and elastase (ELANE) genes in American, African, European, and Asian populations (Vargas-Alarcón et al., 2020).

Furthermore, it was observed that the percentage of invariant residues was inversely proportional to the FD of NFTS of any SARS-CoV-2 protein. In other words, the FD near 2 of an NFTS associated witha SARS-CoV-2 protein signifies wide variations of the protein. So, the widest variations in each SARS-CoV-2 protein variant were observed in the USA and UK since October 2020. On the contrary, the second wave of the COVID-19 infection rate has seen a slight upward trend in Brazil and India, quite significantly. The emerging SARS-CoV-2 strains are expected to contain prevalent mutations in the ORF3a, M, and ORF8 proteins (Table 6). We have found that SARS-CoV-2 proteins were evolving at a high rate considering global data mutations detected in the majority of the available amino acid locations. It was found that the most prominent variants across the globe have one mutation in all proteins. Fractal analysis of NFTS of each protein showed a periodically aperiodic emergence of dominant variants for SARS-CoV-2 protein mutations across various countries. This hidden periodically aperiodic phenomenon of mutations has a strong similarity with forest-fire models (Grassberger, 1993; Mukaka, 2012), where periodic emergence of fire happens after a brief period of non-activity (Drossel and Schwabl, 1992). This possibly distances SARS-CoV-2 from other beta-coronaviruses. Furthermore, it was noticed that there is a negative correlation between the frequency of a variant and when it was first recorded in country-wise data. This signifies that older variants are by and large more dominant, although there were many notable exceptions, and certain variants have quickly become dominant in the last couple of months.

The origin of SARS-CoV-2 has continued to receive much attention even after almost two years since the outbreak. Recombinant events and inevitable mutations keep developing during the pandemic (Danchin and Timmis, 2020). As a consequence of Muller's ratchet, mutations will deplete the genome in its cytosine content, which leads to attenuation of the virus (Muller, 1964). However, this process is counteracted by recombination, which can occur when different coronaviruses or mutants/variants co-infect the same host, creating an enormous evolutionary landscape for viruses to explore (Joffrin et al., 2020). It has been postulated that SARS-CoV-2 might have been active well before the outbreak in Wuhan and either through zoonosis or accidental release from a laboratory, mutations in the SARS-CoV-2 S protein dramatically increased its transmissibility leading to the COVID-19 pandemic (Platto et al., 2021). For this reason and to limit the emergence and impact on the pandemic, it is of utmost importance to evaluate SARS-CoV-2 mutants/variants in the light of the origin of SARS-CoV-2 (Otto et al., 2021).

In conclusion, the periodically aperiodic nature of the spread of mutant SARS-CoV-2 proteins is a unique feature of SARS-CoV-2 among beta-coronaviruses, which can contribute to the identification of the origin of SARS-CoV-2.

Authors' contributions

SSH designed the study. SSH, PB, KL, and VNU contributed to the implementation of the research, to the analysis of the results. SSH and PB wrote the initial draft of the manuscript. SSH, KL, VNU, EMR, PPC, ASA, GKA, AAAA, TMAE, DB, PA, KT, MT, AL, GC, and SPS reviewed and edited. GP, BDU, WBC, and NGB provided constructive reviews and suggestions. All authors read the final version and approve.

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.

Acknowledgements

We gratefully acknowledge the authors from laboratories responsible for obtaining the specimens and submitting sequence data, shared via the GISAID Initiative, on which this research is based.

References

Asselta, R., Paraboschi, E.M., Mantovani, A., Duga, S., 2020. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. Aging (N Y) 12, 10087.
Bianchi, M., Borsetti, A., Ciccozzi, M., Pascarella, S., 2021. SARS-CoV-2 ORF3a: mutability and function. Int. J. Biol. Macromol. 170, 820–826.
Chen, J., Wang, R., Wang, M., Wei, G.W., 2020. Mutations strengthened SARS-CoV-2 infectivity. J. Mol. Biol. 432, 5212–5226.
Danchin, A., Timmis, K., 2020. SARS-CoV-2 variants: relevance for symptom granularity, epidemiology, immunity (herd, vaccines), virus origin and containment? Environ. Microbiol. 22, 2001–2006.
Domingo, J.L., 2021a. SARS-CoV-2/COVID-19: Natural or Laboratory Origin? Environ. Res. 201, 111542.
Domingo, J.L., 2021b. What We Know and What We Need to Know about the Origin of SARS-CoV-2. Environ Res. 200, 111785.
Domingo, J.L., Marqués, M., Rovira, J., 2020. Influence of airborne transmission of SARS-CoV-2 on COVID-19 pandemic. a review. Environ. Res. 188, 109861.
van Dorp, L., Richard, D., Tan, C.C., Shaw, L.P., Acman, M., Balloux, F., 2020. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. Nat. Commun. 11, 1–8.
Drossel, B., Schwabl, F., 1992. Self-organized critical forest-fire model. Phys. Rev. Lett. 69, 1629.
Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D.d.S., Mishra, S., Crispim, M.A., Sales, F.C., Hawryluk, I., McCrone, J.T., et al., 2021. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science 372, 815–821.
Gelly, J.C., Joseph, A.P., Srinivasan, N., de Brevern, A.G., 2011. iPBA: A tool for protein structure comparison using sequence alignment strategies. Nucleic Acids Res. 39, W18–W23.
Gómez-Carballa, A., Bello, X., Pardo-Seco, J., Martinón-Torres, F., Salas, A., 2020. Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Res. 30, 1434–1448.
Grassberger, P., 1993. On a self-organized critical forest-fire model. J. Phys. Math. Gen. 26, 2081.
Gupta, R., Charron, J., Stenger, C.L., Painter, J., Steward, H., Cook, T.W., Faber, W., Frisch, A., Lind, E., Bauss, J., et al., 2020. SARS-CoV-2 (COVID-19) structural and evolutionary dynamicome: insights into functional evolution and human genomics. J. Biol. Chem. 295, 11742–11753.
Hassan, S.S., Pal Choudhury, P., Basu, P., Jana, S.S., 2020. Molecular conservation and differential mutation on ORF3a gene in Indian SARS-CoV-2 genomes. Genomics 112, 3226–3237.
Hassan, S.S., Kodakandla, V., Redwan, E.M., Lundstrom, K., Choudhury, P.P., Mohamed Abd El-Aziz, T., Takayama, K., Kandimalla, R., Lal, A., Serrano-Aroca, Á., et al.,

2021a. An Issue of Concern: Unique Truncated ORF8 Protein Variants of SARS-CoV-2. bioRxiv. https://doi.org/10.1101/2021.05.25.445557.
Hassan, S.S., Kodakandla, V., Redwan, E.M., Lundstrom, K., Pal Choudhury, P., Serrano-Aroca, Á., Kumar Azad, G., Aljabali, A.A.A., Palú, G., Mohamed Abd El-Aziz, T, et al., 2021b. Non-uniform aspects of SARS-CoV-2 intraspecies evolution reopen questions on its origin. Preprints. https://doi.org/10.20944/preprints202106.0472.v1, 2021060472.
Health, T.L.P., 2020. COVID-19 puts societies to the test. The Lancet. Publ. Health 5, e235.
Hollis, N.D., Li, W., Van Dyke, M.E., Njie, G.J., Scobie, H.M., Parker, E.M., Penman-Aguilar, A., Clarke, K.E., 2021. Racial and ethnic disparities in incidence of SARS-CoV-2 infection, 22 US states and DC, January 1–October 1, 2020. Emerg. Infect. Dis. 27, 1477.
Joffrin, L., Goodman, S.M., Wilkinson, D.A., Ramasindrazana, B., Lagadec, E., Gomard, Y., Le Minter, G., Dos Santos, A., Schoeman, M.C., Sookhareea, R., et al., 2020. Bat coronavirus phylogeography in the Western Indian Ocean. Sci. Rep. 10, 1–11.
Khan, M.I., Khan, Z.A., Baig, M.H., Ahmad, I., Farouk, A.E., Song, Y.G., Dong, J.J., 2020. Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: an in silico insight. PloS One 15, e0238344.
Ko, K., Nagashima, S., E, B., Ouoba, S., Akita, T., Sugiyama, A., Ohisa, M., Sakaguchi, T., Tahara, H., Ohge, H., et al., 2021. Molecular characterization and the mutation pattern of SARS-CoV-2 during first and second wave outbreaks in Hiroshima, Japan. PloS One 16, e0246383.
Kumar, R., Verma, H., Singhvi, N., Sood, U., Gupta, V., Singh, M., Kumari, R., Hira, P., Nagar, S., Talwar, C., et al., 2020. Comparative genomic analysis of rapidly evolving SARS-CoV-2 reveals mosaic pattern of phylogeographical distribution. mSystems 5 e00505–e00520.
Kupferschmidt, K., 2020. The pandemic virus is slowly mutating. but does it matter?.
Lee Rodgers, J., Nicewander, W.A., 1988. Thirteen ways to look at the correlation coefficient. Am. Statistician 42, 59–66.
Liu, P., Chen, W., Chen, J.P., 2019. Viral metagenomics revealed Sendai virus and Coronavirus infection of Malayan pangolins (Manis javanica). Viruses 11, 979.
Liu, P., Jiang, J.Z., Wan, X.F., Hua, Y., Li, L., Zhou, J., Wang, X., Hou, F., Chen, J., Zou, J., et al., 2020. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLoS Pathog. 16, e1008421.
Lv, L., Li, G., Chen, J., Liang, X., Li, Y., 2020. Comparative Genomic Analyses Reveal Specific Mutation Pattern between Human Coronavirus SARS-CoV-2 and Bat-CoV RaTG13, 11. Front. Microbiol. 11, 584717.
McCoy, J., Wambier, C.G., Vano-Galvan, S., Shapiro, J., Sinclair, R., Ramos, P.M., Washenik, K., Andrade, M., Herrera, S., Goren, A., 2020. Racial variations in COVID-19 deaths may be due to androgen receptor genetic variants associated with prostate cancer and androgenetic alopecia. Are anti-androgens a potential treatment for COVID-19? J. Cosmet. Dermatol. 19, 1542–1543.
Mercatelli, D., Giorgi, F.M., 2020. Geographic and genomic distribution of SARS-CoV-2 mutations. Front. Microbiol. 11, 1800.
Mukaka, M.M., 2012. A guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. 24, 69–71.
Muller, H.J., 1964. The relation of recombination to mutational advance. Mutat. Res. Fund Mol. Mech. Mutagen 1, 2–9.
Nadeau, S.A., Vaughan, T.G., Scire, J., Huisman, J.S., Stadler, T., 2021. The origin and early spread of SARS-CoV-2 in Europe. Proc. Natl. Acad. Sci. Unit. States Am. 118, e2012008118 https://doi.org/10.1073/pnas.2012008118.
Ogando, N.S., Zevenhoven-Dobbe, J.C., van der Meer, Y., Bredenbeek, P.J., Posthuma, C. C., Snijder, E.J., 2020. The enzymatic activity of the NSP14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. J. Virol. 94 e01246–20.
Otto, S.P., Day, T., Arino, J., Colijn, C., Dushoff, J., Li, M., Mechai, S., Van Domselaar, G., Wu, J., Earn, D.J., et al., 2021. The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. Curr. Biol. 31, R918–R929.
Pedersen, S.F., Ho, Y.C., 2020. SARS-CoV-2: A storm is raging. J. Clin. Invest. 130, 2202–2205.
Pilgrim, I., Taylor, R., 2018. Fractal Analysis of Time-Series Data Sets: Methods and Challenges, in: Fractal Analysis. IntechOpen.
Pipes, L., Wang, H., Huelsenbeck, J.P., Nielsen, R., 2021. Assessing uncertainty in the rooting of the SARS-CoV-2 phylogeny. Mol. Biol. Evol. 38, 1537–1543.
Platto, S., Wang, Y., Zhou, J., Carafoli, E., 2021. History of the COVID-19 pandemic: origin, explosion, worldwide spreading. Biochem. Biophys. Res. Commun. 538, 14–23.
Romano, M., Ruggiero, A., Squeglia, F., Maga, G., Berisio, R., 2020. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. Cells 9, 1267.
Sallard, E., Halloy, J., Casane, D., Decroly, E., van Helden, J., 2021. Tracing the origins of SARS-CoV-2 in coronavirus phylogenies: a review. Environ. Chem. Lett. 1–17.
Sánchez-Granero, M., Fernández-Martinez, M., Trinidad-Segovia, J., 2012. Introducing fractal dimension algorithms to calculate the hurst exponent of financial time series. The European Physical Journal B 85, 1–13.
Setti, L., Passarini, F., De Gennaro, G., Barbieri, P., Perrone, M.G., Borelli, M., Palmisani, J., Di Gilio, A., Torboli, V., Fontana, F., et al., 2020. SARS-CoV-2 RNA found on particulate matter of Bergamo in northern Italy: first evidence. Environ. Res. 188, 109754.
Seyran, M., Pizzol, D., Adadi, P., El-Aziz, T.M., Hassan, S.S., Soares, A., Kandimalla, R., Lundstrom, K., Tambuwala, M., Aljabali, A.A., et al., 2021. Questions concerning the proximal origin of SARS-CoV-2. J. Med. Virol. 93, 1204.
Shoily, S.S., Ahsan, T., Fatema, K., Sajib, A.A., 2021. Disparities in COVID-19 severities and casualties across ethnic groups around the globe and patterns of ACE2 and PIR variants. Infect. Genet. Evol. 92, 104888.
Shu, Y., McCauley, J., 2017. Gisaid: global initiative on sharing all influenza data–from vision to reality. Euro Surveill. 22, 30494.
Srivastava, A., Bandopadhyay, A., Das, D., Pandey, R.K., Singh, V., Khanam, N., Srivastava, N., Singh, P.P., Dubey, P.K., Pathak, A., et al., 2020. Genetic association of ACE2 rs2285666 polymorphism with COVID-19 spatial distribution in India. Front. Genet. 11, 1163.
Sze, S., Pan, D., Nevill, C.R., Gray, L.J., Martin, C.A., Nazareth, J., Minhas, J.S., Divall, P., Khunti, K., Abrams, K.R., et al., 2020. Ethnicity and Clinical Outcomes in COVID-19: a Systematic Review and Meta-Analysis. EClinicalMedicine, p. 100630.
Tapper, E.B., Asrani, S.K., 2020. COVID-19 pandemic will have a long-lasting impact on the quality of cirrhosis care. J. Hepatol. 73, 441–445.
Vahidy, F.S., Nicolas, J.C., Meeks, J.R., Khan, O., Pan, A., Jones, S.L., Masud, F., Sostman, H.D., Phillips, R., Andrieni, J.D., et al., 2020. Racial and ethnic disparities in SARS-CoV-2 pandemic: analysis of a COVID-19 observational registry for a diverse us metropolitan population. BMJ open 10, e039849.
Vargas-Alarcón, G., Posadas-Sánchez, R., Ramırez-Bello, J., 2020. Variability in genes related to sars-cov-2 entry into host cells (ace2, tmprss2, tmprss11a, elane, and ctsl) and its potential use in association studies. Life Sci. 260, 118313.
Wang, Y., Wang, D., Zhang, L., Sun, W., Zhang, Z., Chen, W., Zhu, A., Huang, Y., Xiao, F., Yao, J., et al., 2021. Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients. Genome Med. 13, 1–13.
Williams, T.C., Burgers, W.A., 2021. SARS-CoV-2 evolution and vaccines: cause for concern? The Lancet Respiratory Medicine 9, 333–335.
Yuan, M., Huang, D., Lee, C.C.D., Wu, N.C., Jackson, A.M., Zhu, X., Liu, H., Peng, L., van Gils, M.J., Sanders, R.W., et al., 2021. Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. Science 373, 818–823.
Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273.