



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.



# Geographical distribution of host's specific SARS-CoV-2 mutations in the early phase of the COVID-19 pandemic

Mohammad Khalid<sup>a,\*</sup>, David Murphy<sup>b</sup>, Maryam Shoai<sup>c</sup>,  
Jonahunnatha Nesson George-William<sup>d</sup>, Yousef Al-ebini<sup>a</sup>

<sup>a</sup> Department of Pharmaceutics, College of Pharmacy, King Khalid University, Asir-Abha 61421, Saudi Arabia

<sup>b</sup> Department of Clinical and Movement Neurosciences, UCL Institute of Neurology, Queen Square, London, United Kingdom

<sup>c</sup> Department of Neurodegenerative Diseases, UCL Institute of Neurology, London, United Kingdom

<sup>d</sup> Center for Advanced Studies and Technology (CAST), G. d'Annunzio University, 66100 Chieti, Italy

## ARTICLE INFO

Edited by: Yoshiyuki Suzuki

### Keywords:

Mutation  
SARS-CoV-2  
NSP6  
ORF-8  
Nucleotide sequences

## ABSTRACT

**Purpose:** To assess, if the SARS-CoV-2 mutate in a similar pattern globally or has a specific pattern in any given population.

**Results:** We report, the insertion of TTT at 11085, which adds an extra amino acid, F to the NSP6 at amino acid position 38. The highest occurrence of TTT insertion at 11,085 position was found in UK derived samples (65.97%). The second and third highest occurrence of the mutation were found in Australia (8.3%) and USA (4.16%) derived samples, respectively.

Another important discovery of this study is the C27945T mutation, which translates into the termination of ORF-8 after 17 amino acids, reveals that the SARS-CoV-2 can replicate without the intact ORF-8 protein. We found that the 97% of C27945T mutation of global occurrence, occurred in Europe and the USA derived samples. **Conclusions:** Two of the reported mutations (11085TTT insertion and C27945T nonsense), which seemed to reduce Type I interferon response are linked to specific geographical locations of the host and implicate region-specific mutations in the virus. The findings of this study signify that SARS-CoV-2 has the potential to adapt differently to different populations.

## 1. Introduction

The coronaviruses belong to the order Nidovirales, family Coronaviridae, and subfamily Coronavirinae (Cui et al., 2019). The COVID-19 pandemic is caused by the novel coronavirus, named as 2019-nCoV or SARS-CoV-2, belonging to the Betacoronavirus genus of the subfamily Coronavirinae. The SARS-CoV-2 virus contains a positive-sense single-stranded RNA (+ssRNA) genome with a size of around 29.9 kilobases (Chan et al., 2020). The genomes of coronaviruses are among the longest RNA viruses (Gorbalenya et al., 2006). Although coronavirus' RNA-dependent RNA polymerase (RdRp) have a proofreading function, single-stranded RNA genomes have a higher mutation rate than double-stranded DNA viruses (Peck and Lauring, 2018 May 2). Another cause of mutation of the viral genome is the driving force to evade the host immune system defences. These factors make mutant variants of various

viral proteins, which help the virus in evading the defences. Consequently, the fitting variants prevail in any geographical region/population over time.

Mutational changes in the viral genome may lead to sequence changes in untranslated regions (UTRs) as well as changes in amino acid sequences of structural and functional proteins of the viruses. Being the RNA virus, the SARS-CoV-2 is mutating; several rare and common mutations in UTR/ORFs in the genome have been already reported (Phan, 2020 Jul; Tang et al., 2020 Jun 1; Wang et al., 2020; Stefanelli et al., 2020; Kim et al., 2020; Pachetti et al., 2020 Dec 22). The D614G mutation in spike protein has got special attention among the SARS-CoV-2 virologist for becoming dominant around the world (Korber et al., 2020) and the mutation also correlated with a higher mortality rate in COVID-19 patients (Toyoshima et al., 2020 Dec 22). It is very likely that the SARS-CoV-2 may have more mutations like D614G, which may affect

**Abbreviations:** GISAID, Global initiative on sharing all influenza data; IRF3, Interferon regulatory factor 3; NCBI, National Center for Biotechnology Information; NSP6, Non-structural protein 6; ORF-8, Open Reading Frame-8; RdRp, RNA dependent RNA polymerase; TBK1, TANK binding kinase 1; UTRs, Untranslated regions.

\* Corresponding author at: Department of Pharmaceutics, College of Pharmacy, King Khalid University, Abha-61421 Asir, Saudi Arabia.

E-mail addresses: [mmabood@kku.edu.sa](mailto:mmabood@kku.edu.sa), [1khalidmd@gmail.com](mailto:1khalidmd@gmail.com) (M. Khalid).

<https://doi.org/10.1016/j.gene.2022.147020>

Received 4 June 2022; Received in revised form 24 October 2022; Accepted 25 October 2022

Available online 4 November 2022

0378-1119/© 2022 Elsevier B.V. All rights reserved.

the viral infectivity and clinical outcome of the infection and hence, this study aims to explore such mutations more thoroughly.

Our aim was to analyse the SARS-CoV-2 genome sequences derived from patients of all six continents and to investigate whether the virus mutate in similar pattern globally or has specific pattern in any given populations or hosts. In the present, study we analysed 15,120 full length SARS-CoV-2 genomes, derived from symptomatic/asymptomatic COVID-19 patients from all six continents, submitted to NCBI Database (<https://www.ncbi.nlm.nih.gov>). We further investigated an additional 1,000 SARS-CoV-2 genome sequences, derived from Australia, China, India, Saudi Arabia (KSA), Egypt, Italy, Spain, USA, Mexico and Brazil submitted to GISAID (<https://www.gisaid.org>).

## 2. Methods

### 2.1. SARS-CoV-2 genome sequences

We analysed 15,120 full-length SARS-CoV-2 genomes (global set), derived from symptomatic/asymptomatic COVID-19 patients. These sequences were submitted to NCBI database from all six continents. We downloaded all the SARS-CoV-2 genome sequences available from the NCBI database as on June 25, 2020.

We also investigated another 1,000 SARS-CoV-2 genome sequences (country set) from symptomatic/asymptomatic COVID-19 patients from 10 countries (Australia, China, India, KSA, Egypt, Italy, Spain, USA, Mexico and Brazil, covering all six continents) as large number of SARS-CoV-2 infection cases have been reported from these countries. The sequences were downloaded from GISAID and 100 full-length SARS-CoV-2 genome sequences were randomly selected from each country.

### 2.2. Mutational analysis of the SARS-CoV-2 genome sequences

The 15,120 full-length SARS-CoV-2 genome sequences were processed using FASTA\_Unique\_Sequences\_1.0 tool (<https://www.ncbi>). Next, we used a sequencing tool, Minimap2 (Li, 2018 Sep 15), which is designed for long-read sequences with a larger number of samples to be aligned. Using Minimap2, once aligned with a reference sequence (accession number NC\_045512.2) (WuF and YuB, , 2020), the duplicate sequences were removed more aggressively and that gave us 400 genome sequences with mutations (Workflow: Step A). Although the MiniMap2 is suited for longer sequences and large number of samples, it has the disadvantage of showing the most common mutations only and hence we had to utilize another approach to find infrequent mutations.

To analyse the 1000 full-length SARS-CoV-2 genome sequences, derived from the aforementioned countries, we have designed a software tool, for fragmentation of nucleotide sequences. This tool helped us to fragment the SARS-CoV-2 genome sequences into 5000 base sections (Khalid et al., 2022). We have used MultAlin online tool (<http://multalin>) for alignment of fragmented genome sequences with the reference viral strain sequence mentioned above. We have taken care of not to miss part of the genome sequences by having 100 bases long overlapping sequences in each fragment. We considered variation as true mutation if variations from reference sequence were identified in more than one instances. The same practice was applied with the genome sequence samples of all six countries (Workflow: Step B). The results were validated by the MiniMap2 too as well as with published data.

### 2.3. The occurrence of 11085TTT insertion and C27945T mutations in the SARS-CoV-2 genome

We downloaded all the SARS-CoV-2 genome sequences from GISAID database available on September 2, 2020, that provided us 93,265 genome sequences. We focussed our analysis for occurrence of 11085TTT insertion that translate to Phenylalanine (F) amino acid at the position 38 in non-structural protein 6 (NSP6). Additionally, C27945T mutation, which translates to stop codon after 17 amino acids in ORF-8

only from these 93,265 genome sequences were analysed (Workflow: Step C).

## 3. Results

### 3.1. Workflow of mutation analysis

We have used the following workflow for the mutation analysis (see Fig. 1).

### 3.2. Common SARS-CoV-2 mutations across the globe

We analysed total 15,120 full-length SARS-CoV-2 genome sequences (global set), available on the NCBI database as on 25 June 2020, investigating mutations occurrence since COVID-19 outbreak. The alignment was made with the viral reference strain (accession number NC\_045512.2) of the virus. We detected 11 synonymous and non-synonymous mutations from the global set of data (Table 1 and Supplementary Fig. 1), these mutations were previously reported (Phan, 2020 Jul; Tang et al., 2020 Jun 1; Wang et al., 2020; Stefanelli et al., 2020; Pachetti et al., 2020 Dec 22). From these 11 mutations the C241T mutation is occurring in 5'-UTR regions of the virus genome. The mutation was identified in quite significant number of samples (over 85% in all countries except China USA and Spain where it was 3%, 40% and 56%, respectively). Three of the 11 mutations are synonymous mutations. The remaining 8 lead to protein change in open reading frames 1a, 1b, S, 3a and 8 (Table 1, Supplementary Fig. 1).

### 3.3. Unique mutation in the SARS-CoV-2 samples from each continent

We were also interested in mutations occurrence in a minute percentage of samples globally. We have chosen Australia as representative of Oceania continent, China, India and KSA as representative of Asia, Egypt representing Africa, Italy and Spain as representative of Europe, USA for North America, Mexico for Central America and Brazil as a representative of South American continent. The chosen countries were having quite notable numbers of SARS-CoV-2 infection cases. We used the GISAID database to download the 100 full-length genome sequences derived from the above-mentioned countries, selected at random. Alignment of these samples to reference strain showed us a total of 313 mutations in SARS-CoV-2 genome in country set of data (Supplementary Table 1). Many of these mutations are rare but a considerable number of mutations are common in samples derived from all six continents (Table 2). All the 11 mutations found in the global set of data were also included in the 313 mutations, obtained from the country set of data.

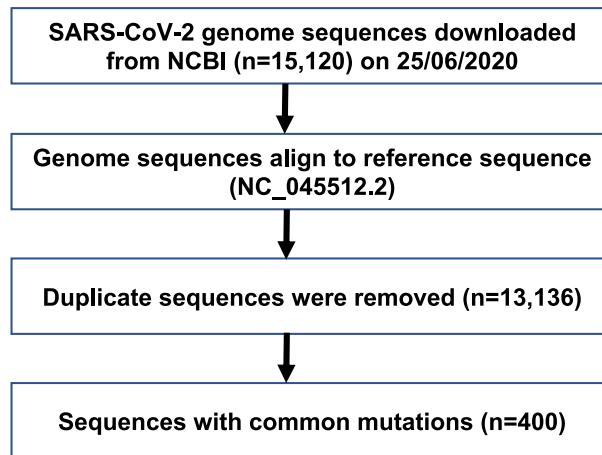
### 3.4. Country specific mutations

The 313 mutations were found at the nucleotide level in most of the ORFs, except in ORF 7b, 10 and 3'-UTR, (Fig. 2). The highest mutation occurrence (78 mutations; 24.9%) were detected in Australian samples, whilst the lowest mutation occurrence compared to reference strain, were found in Brazilian samples (28 mutations; 8.9%) (Fig. 2 and Supplementary Table 1).

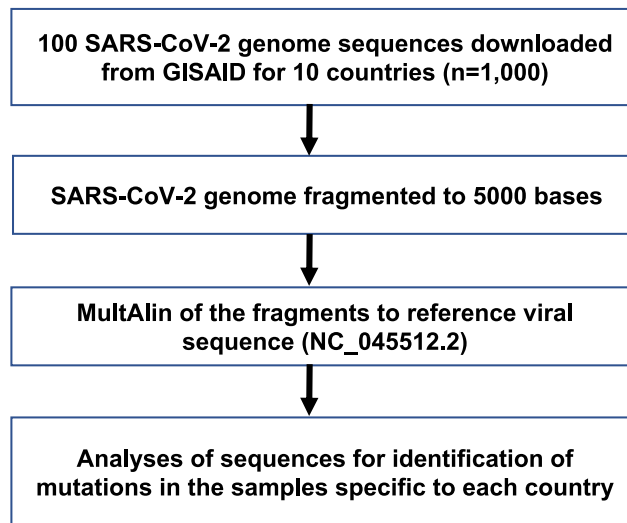
### 3.5. Mutations at amino acid level

All 313 mutations were analysed, for their corresponding amino acid changes, as shown in (Fig. 3). The highest non-synonymous mutations (43 mutations; 13.7%), we detected in Australian samples, which corresponds to the highest nucleotide mutations (78 mutations). While the lowest non-synonymous mutations were found in KSA derived samples (17 mutations; 5.4%).

**Step A:** The mutation workflow analysis of SARS-CoV-2 genome sequences obtained from NCBI database (global set on 25/06/2020) to identify common mutations globally.



**Step B:** The mutation workflow analysis of SARS-CoV-2 genome sequences obtained from the GISAID database (country set, 100 from each country as mentioned earlier).



**Step C:** The occurrence of the 11085TTT insertion and the C27945T mutations globally (on 02/09/2020).

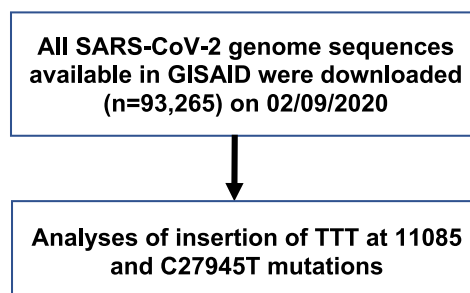


Fig. 1. Workflow of identification of mutations in SARS-CoV-2 genome sequences.

### 3.6. Most of the nucleotide mutations are non-synonymous mutations

We noticed that non-synonymous mutations are higher than synonymous mutations in majority of ORFs (Fig. 4). We discovered 2% of the

genome sequences in Australian samples have insertion of TTT at position 11,085 which lies in ORF-1a. The mutation adds one amino acid, F at the 3607 position in the ORF-1a, more precisely, the mutation adds F amino acid at the position 38 in non-structural protein 6 (NSP6), which

**Table 1**  
Common SARS-CoV-2 Mutations Across the Globe.

S. No.	Nucleotide	Occurrence	ORF	Amino Acid
1	C241T	61%	5'-UTR	N/A
2	C1059T	28%	1a	T265I
3	C3037T	52%	1a	Silent
4	C8782T	31%	1a	Silent
5	C14408T	52%	1b	P314L
6	C17747T	26%	1b	P1427L
7	A17858G	26%	1b	Y1464C
8	C18060T	26%	1b	Silent
9	A23403G	37%	S	D614G
10	G25563T	38%	3a	Q57H
11	T28144C	25%	8	L84S

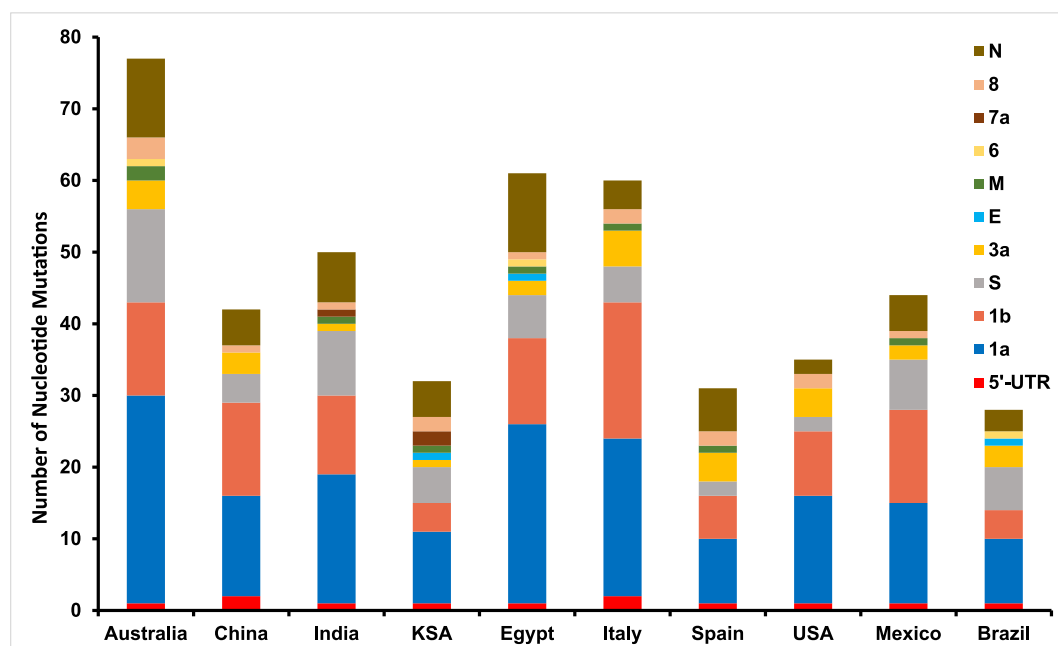
expresses as part of ORF-1a. Another independent mutation that interests us at the position C27945T in 2% Italian samples, which translate to non-sense codon in the ORF-8. The mutation creates premature stop codon and does not allow the ORF-8 to express after the 17 amino acid although it is 121 amino acid long ORF, as per the reference strain (Supplementary Fig. 3).

**3.7. The occurrence of nsp6 insertion mutation predominantly in UK derived samples**

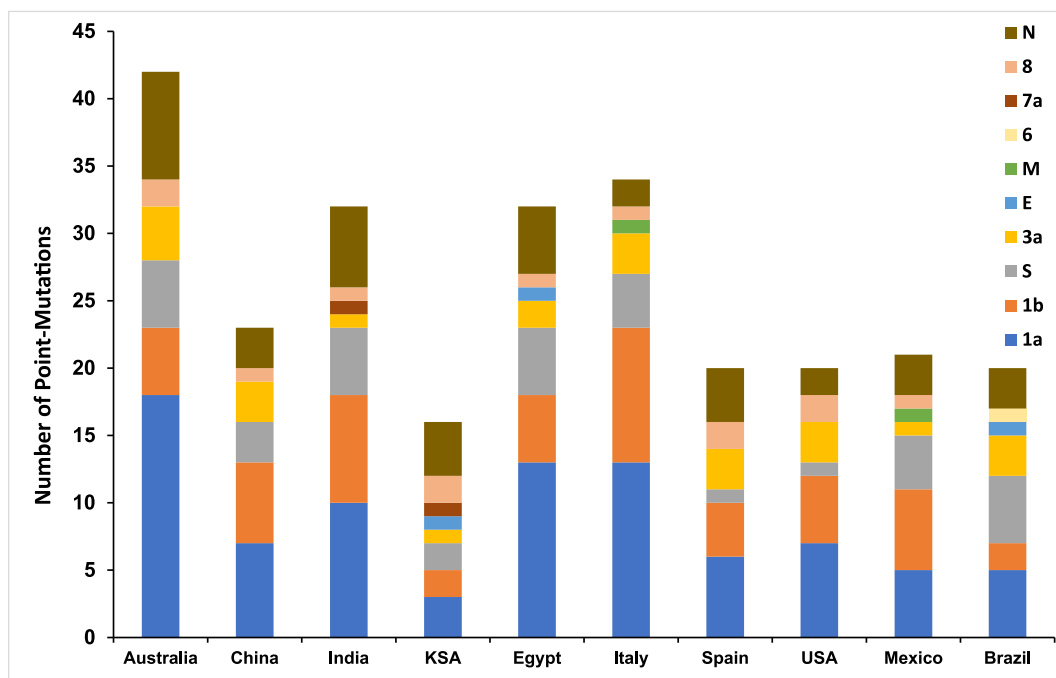
We analyzed the occurrence of 11085TTT insertion mutation in the nsp6 in the GISAID. From the analysis of 93,265 SARS-CoV-2 genome sequences, we found the insertion mutation in the nsp6, only in 288 samples. Remarkably, out of the 288 samples, 190 (65.97%) samples were derived from UK patients. Australia derived samples are the second-largest (8.3%) after the UK, with the insertion mutation (Fig. 5,

**Table 2**  
The prevalence of mutations in samples derived from representative countries of all six continents.

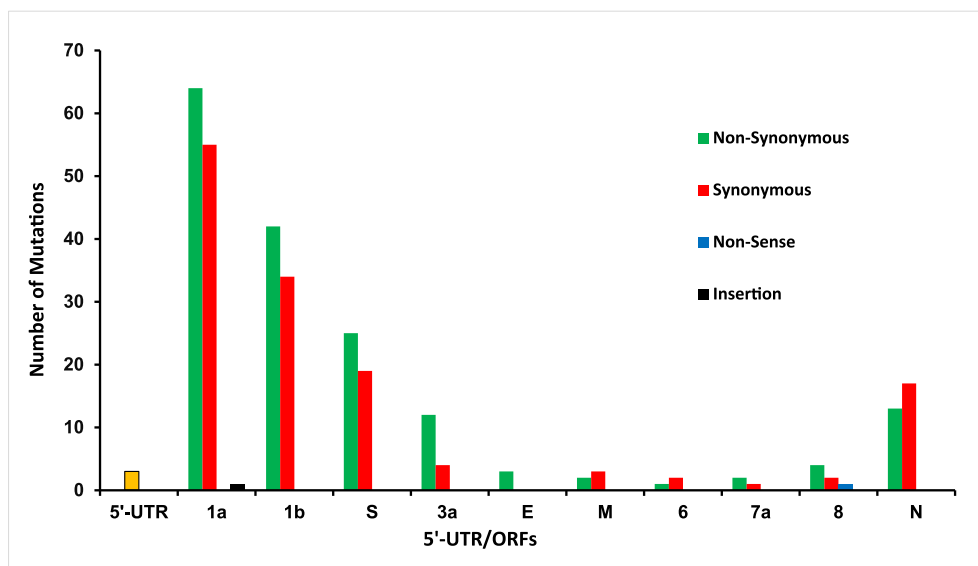
S. No.	Mutation	Australia	China	India	KSA	Egypt	Italy	Spain	USA	Mexico	Brazil	ORF	Amino Acid
1	C241T	85%	3%	94%	89%	90%	100%	56%	40%	86%	100%	5'-UTR	—
2	C1059T	13%			2%			4%	28%	12%	15%	1a	T265I
3	C3037T	90%	3%	98%	90%	97%	100%	57%	42%	86%	100%	1a	Silent
4	C8782T	5%	48%	4%	8%	2%		42%	33%	12%		1a	Silent
5	G11083T	9%	6%				15%	5%	5%	4%	15%	1a	L3606F
6	C14408T	88%	3%	90%	92%	65%	100%	56%	40%	86%	100%	1b	P314L
7	C14805T	6%						31%	2%		15%	1b	Silent
8	C17747T								35%	7%		1b	P1427L
9	A17858G								35%	7%		1b	Y1464C
10	C18060T		3%						35%	9%		1b	Silent
11	C18877T	5%		50%	61%	25%			1%	2%		1b	Silent
12	A20268G		2%				7%	40%		55%		1b	Silent
13	C22444T	5%		40%	11%							S	Silent
14	A23403G	80%	3%	94%	90%	98%	100%	60%	40%	86%	100%	S	D614G
15	G25563T	20%		50%	58%	70%		4%	35%	14%		3a	Q57H
16	G26144T	2%	4%				15%	5%	2%		15%	3a	G251V
17	C26735T	5%		50%	60%	3%						M	Silent
18	T28144C	5%	48%	3%	8%	4%		42%	33%	12%		8	L84S
19	C28854T	3%		40%	9%	2%				41%		N	S194L
20	GG28881AA	60%	1%	1%	23%	21%	40%	6%	5%	16%	99%	N	R203K
21	G28883C	60%	1%	1%	23%	21%	40%	6%	5%	16%	99%	N	G204R



**Fig. 2. SARS-CoV-2 Mutations in different ORFs** The graph shows the total mutations in the 5'-UTR and various ORFs of the SARS-CoV-2 genome.



**Fig. 3. SARS-CoV-2 Non-Synonymous Mutations in different ORFs** The figure shows the number of non-synonymous mutations in various ORFs of the SARS-CoV-2 genome.



**Fig. 4. The translation of SARS-CoV-2 genomic mutation** The graph shows the number of non-synonymous and synonymous mutations in each ORF. The 5'-UTR does not get translated. The figure shows one insertion mutation in ORF-1a and one non-sense mutation in ORF-8 of the SARS-CoV-2 genome.

Supplementary Table 2).

**3.8. The occurrence of non-sense mutation in the ORF-8 in Europe and USA derived samples**

We also investigated the occurrence of the C27945T mutation that translate to non-sense codon after 17 amino acid in ORF-8. The search resulted to 67 samples, out of 93,265 SARS-CoV-2 genome sequences obtained from the GISAID database, with the C27945T mutation. 97% of these samples were derived from Europe and USA, (Fig. 6A, 6B and Supplementary Table 3). Only two samples were derived from non-western countries; one from Singapore and the another from South Africa.

**4. Discussion**

The RNA genome of SARS-CoV-2 renders the virus prone to mutations. We report novel mutations in the SARS-CoV-2 genome resulting in change in 5'-UTR or translational change, and subsequently altered proteins. The virus appears to be mutating across the globe and some of these mutations are arising independently across various populations and locations.

Many of the mutations found are common in samples, derived from more than one country or continent, as shown in Table 2 and Supplementary Table 1. The result in Table 1 indicates the mutations are highly skewed towards C-to-T transitions (>70%), this mutation bias might be caused by host-specific cytidine deaminases, mostly APOBEC3. The

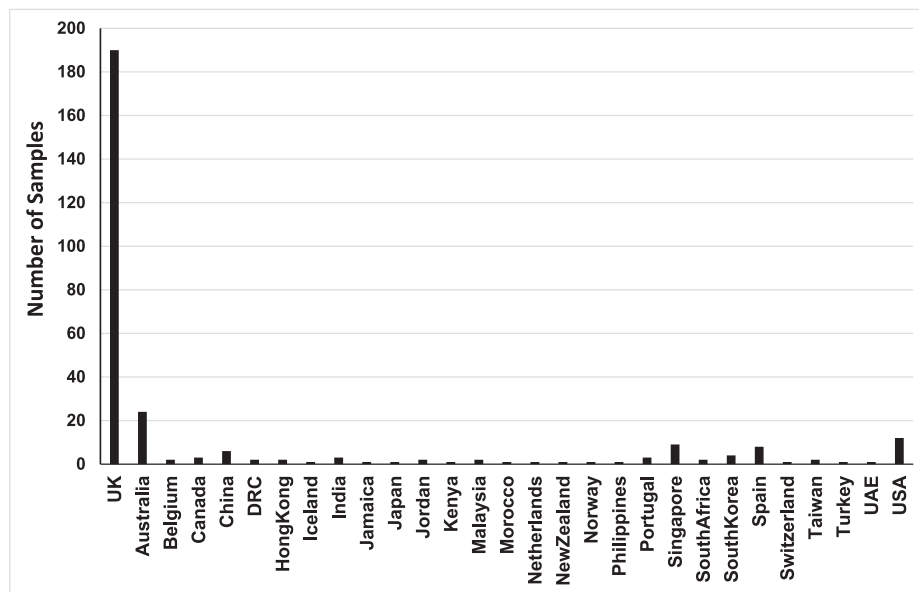


Fig. 5. The occurrence of insertion mutation in SARS-CoV-2 nsp6 The graph shows the occurrence of 11085TTT insertion mutation in SARS-CoV-2 genomes globally..

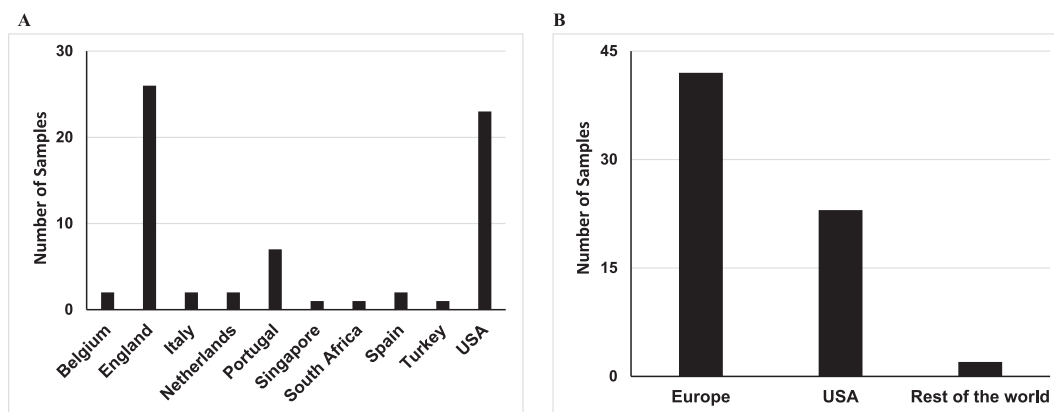


Fig. 6. The occurrence of non-sense mutation in SARS-CoV-2 ORF-8 The graphs show the occurrence of C27945T (non-sense) mutation in ORF-8.

cytidine deaminases systems are directly involved in virus mutagenesis and this study confirms this general trend observed in early studies (Matyášek and Kovařík, 2020; Simmonds, 2020 Jun 24).

Three mutations (C186T, A187G and C241T) were found in 5'-UTR. The C241T mutation is very common and previously reported (Stefanelli et al., 2020), the C186T mutation occur only in 3% of samples derived from main land China and the A187G mutation found in 2% of samples derived from Italy. These two independent mutations were not observed commonly in any other population, suggesting a localised non-advantageous change. The C241T mutation in 5'-UTR, was observed in 7 countries out of 10 with a frequency of more than 85%. The exceptions being China, USA and Spain where the frequency was as low as 3%, 40% and 56%, respectively. The mutation may be providing replication advantage to the virus and the mutant virus is prevailing over time. A detailed study of the C241T mutation is required to shed light on its role in infectivity or pathogenesis. Most of the identified mutations occur in ORF1ab, which is concordant with expectations of a long ORF, however, ORF-8 and ORF-N, despite their relatively small size, have a higher density of mutations as seen in Fig. 4 and Supplementary table 1.

The two mutations (C17747T and A17858G) previously reported only in the USA derived samples (Pachetti et al., 2020 Dec 22), were also found in the current study with a frequency of 35% in USA samples and

7% in Mexican samples. This may be reflective of the geographical proximity of the countries and thus more frequent movement of the population. The other five mutations (C1059T, C3037T, C14408T, C18060T and A23403G), were previously reported as restricted to European derived samples (Pachetti et al., 2020 Dec 22), however, our findings indicate that these mutations are spreading in all six continents in quite significant numbers of samples over time (Table 2 and Supplementary table 1).

The A23403G mutation is translating into D614G in ORF-S, which has been observed globally (Korber et al., 2020) and correlated with mortality rate in COVID-19 patients (Toyoshima et al., 2020 Dec 22). The frequency of the D614G mutation in ORF-S, was found to be more than 80% in 7 out of 10 countries, except China, USA, and Spain, where it is 3%, 40% and 60%, respectively. The frequent simultaneous occurrence of two mutations at the opposite ends of the genome two (C241T and A23403G) is notable. However, in the literature, there is no information on the correlation of these mutation with the mortality rate and human pathology.

Out of the 313 mutations observed, the majority (169 mutations) are non-synonymous mutations. These non-synonymous mutations were found in all ORFs (except ORF 7b, 10 and 3'-UTR) and thus highly likely to affect the infection and pathogenesis of the virus profoundly

(Korber et al., 2020; Toyoshima et al., 2020 Dec 22; Khalid et al., 2012).

The insertion mutation identified in NSP6 protein, which has 7 putative *trans*-membrane helices (Benvenuto et al., 2020) (Supplementary Fig. 2), binds to TANK binding kinase 1 (TBK1) and suppresses the phosphorylation of interferon regulatory factor 3 (IRF3) (Angelini et al., 2013; Xia et al., 2020), thereby, lowering the Type I interferon response; to evade host defences. The insertion of TTT at 11085, which adds an extra amino acid F to the NSP6 at amino acid position 38, occurs mainly in UK (65.97%) derived samples signifying the involvement of host's factor in the mutation. This mutation is also present in samples from Australia (8.3%) and the USA (4.16%), further advocating the involvement of hosts possibly originating from an ancestral UK population, as Australia and USA have significant number of people originating from the UK. The plausible host genetic factors in UK originated samples need to be explored in detail, which may be responsible for the NSP6 insertion mutation.

Lastly, the C27945T mutation, translates into premature termination of ORF-8 (Supplementary Fig. 3). The similar mutations have also been reported from various studies (Khalid et al., 2022; Zinzula, 2020; Pereira, 2020). The ORF-8 protein interacts with several host factors in the lumen of endoplasmic reticulum and also reported to be secreted out of the host's cell (Flower et al., 2021 Jan 12). Unsurprisingly, it plays a role in host's immune response evasion by disrupting the Type I interferon signalling and downregulating MHC-1 (Li et al., 2020; Zhang et al., 2020). These observations are in concordance with published data implicating the interferon deficiency in severe COVID-19 phenotype (Meffre and Iwasaki, 2020). The occurrence of premature termination of ORF-8 in 97% of SARS-CoV-2 isolates derived from countries with predominantly the Caucasian population, indicates the striking connection of this mutation with the said population. The role of the host genetic factor (CCR5 delta 32) has been well documented in reduced HIV-1 infection (Samson et al., 1996). It would be unsurprising if the host genetic factors would be found that play role in the infectivity and replication of the SARS-CoV-2.

## 5. Conclusions

Several of the 313 identified mutations were common in majority of the genome samples irrespective of geographical locations, while many other mutations were specific to a particular population. We report a link between reduced interferon response and SARS-CoV-2 mutations. Two of the reported mutations appear highly linked to the location of the virus and implicate region specific mutations as a response to the host. With the emergence of new variants of SARS-CoV-2 and any potential vaccines being rendered ineffective against the new strains, it is imperative that the structure of SARS-CoV-2 protein and genome are studied in depth in order, to enable identification of areas of the genome with a higher susceptibility to mutations.

## 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.

## Data availability

Data will be made available on request.

## Acknowledgment

We would like to extend our thanks to the people at NCBI database and GISAID for making the SARS-CoV-2 genome sequences available. The authors of this manuscript thanks Dr. Ausaf Ahmad for suggestions and critical reading of the manuscript.

Funding.

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups (Project under grant number RGP.2/244/43).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gene.2022.147020>.

## References

- Angelini, M.M., Akhlaghpour, M., Neuman, B.W., Buchmeier, M.J., 2013. Severe acute respiratory syndrome coronavirus non-structural proteins 3, 4, and 6 induce double-membrane vesicles. *MBio.* 4 (4).
- Benvenuto, D., Angeletti, S., Giovanetti, M., et al., 2020. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. *J Infect.* 81 (1) e24-e27.
- Chan, J.F.W., Kok, K.H., Zhu, Z., et al., 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect.* 9 (1), 221–236.
- Cui, J., Li, F., Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192.
- Flower, T.G., Buffalo, C.Z., Hooy, R.M., Allaire, M., Ren, X., Hurley, J.H., 2021. Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. *Proc. Natl. Acad. Sci. USA.* 2021 Jan 12, 118, 2, e2021785118. <http://doi.org/10.1073/pnas.2021785118>. PMID: 33361333; PMCID: PMC7812859.
- Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. Evolving the largest RNA virus genome. *Nidovirales. Virus Res.* 117, 17–37. <http://multalin.toulouse.inra.fr/multalin>. <https://www.gisaid.org>. [https://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html\\_ncbi/html/software/program.html?uid=11](https://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html_ncbi/html/software/program.html?uid=11). <https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/#nucleotide-sequences>.
- Khalid, M., Yu, H., Sauter, D., et al., 2012. Efficient Nef-Mediated Downmodulation of TCR-CD3 and CD28 Is Associated with High CD4+ T Cell Counts in Viremic HIV-2 Infection. *J. Virol.* 86 (9), 4906–4920.
- Khalid, M., Alshishani, A., Al-ebrini, Y., 2022. Genome Similarities between Human-Derived and Mink-Derived SARS-CoV-2 Make Mink a Potential Reservoir of the Virus. *Vaccines* 10, 1352. <https://doi.org/10.3390/vaccines10081352>.
- Kim, J.S., Jang, J.H., Kim, J.M., Chung, Y.S., Yoo, C.K., Han, M.G., 2020. Genome-Wide Identification and Characterization of Point Mutations in the SARS-CoV-2 Genome. *Osong Public Heal Res. Perspect.* 11 (3), 101–111.
- Korber, B., Fischer, W.M., Gnanakaran, S., et al., 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182 (4), 812–827.e19.
- Li, H., 2018 Sep 15. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 34 (18).
- Li, J.Y., Liao, C.H., Wang, Q., et al., 2020. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. *Virus Res.* 286.
- Matyášek, R., Kovařík, A., 2020. Mutation Patterns of Human SARS-CoV-2 and Bat RaTG13 Coronavirus Genomes Are Strongly Biased Towards C>U Transitions, Indicating Rapid Evolution in Their Hosts. *Genes.* 11 (7), 761. <https://doi.org/10.3390/genes11070761>.
- Meffre, E., Iwasaki, A., 2020. Interferon deficiency can lead to severe COVID. *Nature* 587, 374–376.
- Pachetti, M., Marini, B., Benedetti, F., et al., 2020 Dec 22. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.* 18 (1).
- Peck, K.M., Lauring, A.S., 2018 May 2. Complexities of Viral Mutation Rates. *J. Virol.* 92 (14).
- Pereira, F., 2020. Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. *Infect. Genet. Evol.* 85, 104525.
- Phan, T., 2020 Jul. Genetic diversity and evolution of SARS-CoV-2. *Infect. Genet. Evol.* 81.
- Samson, M., Libert, F., Doranz, B.J., et al., 1996. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR5 chemokine receptor gene. *Nature* 382, 722–725.
- Simmonds, P., 2020. Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. *mSphere.* 2020 Jun 24;5(3):e00408-20. <http://doi.org/10.1128/mSphere.00408-20>. PMID: 32581081; PMCID: PMC7316492.
- Stefanelli, P., Faggioni, G., Lo Presti, A., et al., 2020. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: Additional clues on multiple introductions and further circulation in Europe. *Euro Surveill.* 25 (13).
- Tang, X., Wu, C., Li, X., et al., 2020 Jun 1. On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.* 7 (6).
- Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., Kiyotani, K., 2020 Dec 22. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J. Hum. Genet.* 65 (12).
- Wang, C., Liu, Z., Chen, Z., Huang, X., et al., 2020. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J. Med. Virol.* 92 (6), 667–674.



Wu, F., Zhao, S., Yu, B., et al., 2020. A new coronavirus associated with human respiratory disease in China. *Nature*. 579, 265–269.

Xia, H., Cao, Z., Xie, X., et al., 2020. Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep.* 33 (1).

Zhang, Y., Zhang, J., Chen, Y., et al., 2020. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. *bioRxiv* [Internet]. 2020; 2020.05.24. Available from: <https://doi.org/10.1101/2020.05.24.111823>.

Zinzula, L., 2020. Lost in deletion: The enigmatic ORF8 protein of SARS-CoV-2. *Biochem. Biophys. Res. Commun.* 538, 116–124.