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.

Gene Reports 19 (2020) 100682

Contents lists available at ScienceDirect

Gene Reports

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

Genomic characterization of a novel SARS-CoV-2

Rozhgar A. Khailany[a,*], Muhamad Safdar[b], Mehmet Ozaslan[c]

[a] Department of Biology, College of Science, University of Salahaddin-Erbil, Iraq
[b] Department of Breeding and Genetics, Cholistan University of Veterinary & Animal Sciences, Bahawalpur 63100, Pakistan
[c] Department of Biology, Gaziantep University, 27310 Gaziantep, Turkey

ARTICLE INFO

Keywords:
SARS-CoV-2
Genomic characterization
Mutation
COVID-19

ABSTRACT

A new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) associated with human to human transmission and extreme human sickness has been as of late announced from the city of Wuhan in China. Our objectives were to mutation analysis between recently reported genomes at various times and locations and to characterize the genomic structure of SARS-CoV-2 using bioinformatics programs. Information on the variation of viruses is of considerable medical and biological impacts on the prevention, diagnosis, and therapy of infectious diseases. To understand the genomic structure and variations of the SARS-CoV-2. The study analyzed 95 SARS-CoV-2 complete genome sequences available in GenBank, National MicrobiologyData Center (NMDC) and NGDC Genome Warehouse from December-2019 until 05 of April-2020. The genomic signature analysis demonstrates that a strong association between the time of sample collection, location of sample and accumulation of genetic diversity. We found 116 mutations, the three most common mutations were 8782C > T in ORF1ab gene, 28144T > C in ORF8 gene and 29095C > T in the N gene. The mutations might affect the severity and spread of the SARS-CoV-2. The finding heavily supports an intense requirement for additional prompt, inclusive investigations that combine genomic detail, epidemiological information and graph records of the clinical features of patients with COVID-19.

1. Introduction

The current outbreak of coronavirus disease (COVID-19) that was first reported from Wuhan, China, in December 2019. This epidemic had spread to 206 countries and territories around the world and 2 international conveyances with 1,203,459 confirmed cases, including 64,754 deaths, as of April 05, 2020, so the World Health Organization declared it as a Public Health Emergency of worldwide (https://www.worldometers.info/coronavirus/). Similarly, Middle East respiratory syndrome coronavirus (MERS-CoV) had become a worldwide health concern. MERS-CoV originally reported in 2012 (De Wit et al., 2016). It affected more than 2000 people in 27 countries and 4 sub-continents in the Middle East. While the epidemic of SARS affected 26 countries and resulted in more than 8000 cases in 2003 (De Wit et al., 2016). Since then, a small number of cases have occurred as a result of laboratory accidents or, possibly, through animal-to-human transmission (Guangdong, China) (De Wit et al., 2016).

This study analyzed and discussed available published genome until April 05, 2020, for a better understanding of the genomic variation and characterization of a novel coronavirus (COVID-19). This virus is transmitted from person to person via droplet transmission (Li et al., 2020; Ozaslan et al., 2020). Therefore, the virus is spreading easily in overcrowded areas. Most patients experience only mild to moderate symptoms, such as high body temperature in conjunction with some respiratory symptoms such as cough, sore throat, and headache. Some people may have severe symptoms like pneumonia and acute respiratory distress syndrome (Chen et al., 2020). Also, individuals with underlying complications such as heart disease, chronic lung disease, or diabetes potentially display more severe symptoms (Adhikari et al., 2020). Preventive measures such as masks, frequent hand washing, staying home when sick, avoid public contact, and quarantines are being recommended for reducing the transmission. To date, no specific antiviral treatment is proven effective, hence, infected people initially rely on symptomatic treatments that showed encouraging profile for

Abbreviations: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; COVID-19, Coronavirus disease 2019; NGDC, National Genomics Data Center; NMDC, National Microbiology Data Center; WHO, World Health Organization; EMBOSS, The European Molecular Biology Open Software Suite; BLAST, Basic Local Alignment Search Tool; *UTR*, Untranslated region; CDC, Centers of Disease Control and Prevention; MERS, Middle East Respiratory Syndrome; NCBI, National Center for Biotechnology Information; NSP, nonstructural protein; ORF, Open Reading Frame

* Corresponding author at: Department of Biology, College of Science, University of Salahaddin-Erbil, Erbil 44001, Iraq.
E-mail address: rozhgar.mohammed@su.edu.krd (R.A. Khailany).

https://doi.org/10.1016/j.genrep.2020.100682
Received 26 March 2020; Received in revised form 5 April 2020; Accepted 13 April 2020
Available online 16 April 2020

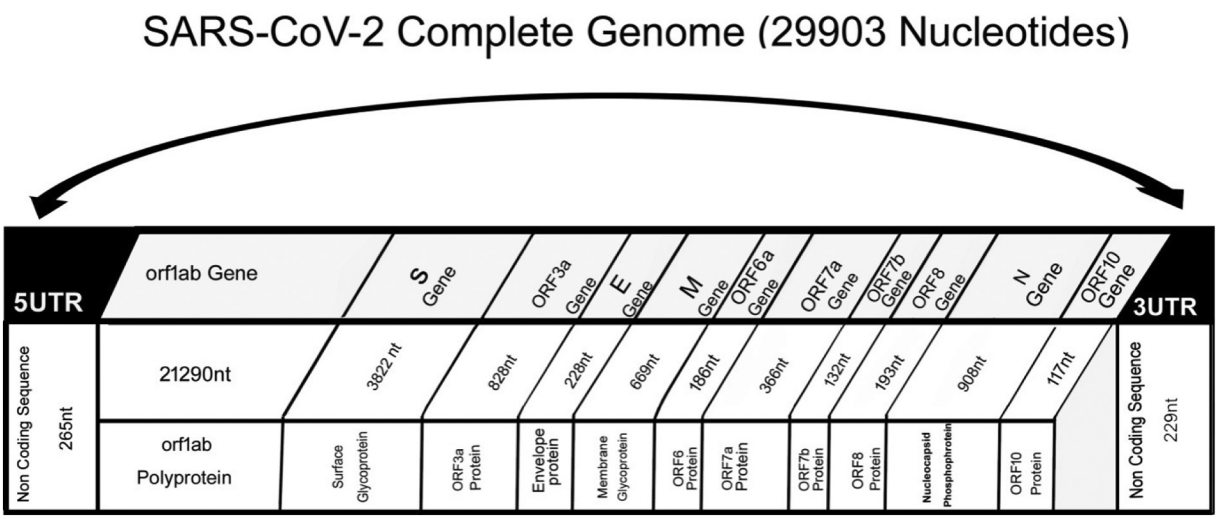

Fig. 1. Structure of the SARS-CoV-2 genome.

blocking the new coronavirus in early clinical trials.

Importantly, the genome size of the SARS-CoV-2 varies from 29.8 kb to 29.9 kb and its genome structure followed the specific gene characteristics to known CoVs; the 5′ more than two-thirds of the genome comprises orf1ab encoding orf1abpolyproteins, while the 3′ one third consists of genes encoding structural proteins including surface (S), envelope (E), membrane (M), and nucleocapsid N proteins (Fig. 1). Additionally, the SARS-CoV-2 contains 6 accessory proteins, encoded by ORF3a, ORF6, ORF7a, ORF7b, and ORF8 genes (Fig. 1) (Li et al., 2005; Oostra et al., 2007).

Recently, the development of high-throughput sequencing has provided datasets of high-quality, complete genome sequences for viral isolates collected in a relatively unbiased manner, regardless of virulence or other unusual characteristics. Analyses of the genome sequence data combined with large-scale antigenic typing have given insights into the pattern of global spread, the genetic diversity during seasonal epidemics, and the dynamics of subtype evolution. SARS-CoV-2 such as the NCBI Severe acute respiratory syndrome coronavirus 2 database (http://www.nhc.gov.cn/jkj/s7915/202001/e4e2d5e6f01147e0a8df3f6701d49f33.shtml) and NGDC Genome Warehouse (bigd.big.ac.cn/gwh/) make the genomic information publicly available, together with epidemiological data for the sequenced isolates. The data sharing requires users to agree to collaborate with, and appropriately credit, all data contributors. A notable success of this initiative has been the contribution of countries, such as China, Philippines, and Japan, etc. which have previously been reticent about placing data in the public domain. The WHO also supports the endeavor of rapid publication of all available sequences for coronaviruses and there is hope that comprehensive submission to public databases will soon become a reality.

The finding heavily supports an intense requirement for additional prompt, inclusive investigations that combine genomic detail, epidemiological information and graph records of the clinical features of patients with COVID-19 (Payne et al., 2018). In the future, mining these resources and establishing a statistical framework based on epidemiological, antigenic, and genetic information could provide further insights into the rules that govern the emergence and establishment of antigenically novel variants and improve the potential for SARS-CoV-2 prevention and control (Ge et al., 2013; Yang et al., 2015). In this study, we investigated the extent of molecular variation between the recently sequenced genomes of SARS-CoV-2.

2. Methodology

We have downloaded 94 publicly available genomes from Genbank up to 12 March. Among 94 genomes, some of the genomes were not used for the analysis due to unusually high variants with gaps. NC_045512 genome sequence was used for reference and the genomic coordinate in this study is based on this reference genome (Lu et al., 2020). Therefore, genomic coordinates must be adjusted to compare with previous studies (Koyama et al., 2020)

Each genome was first aligned to NC_045512 using the EMBOSS needle with a default gap penalty of 10 and an extension penalty of 0.5. Then, differences in comparison with NC_045512 were extracted to create variants (Tables 1 & 2) (Rice et al., 2000). Based on protein annotations, nucleotide level variants were converted into amino acid codon variants for alignments when its location within a gene was identified (Arvestad, 2018). Nucleotide mutations in the genomes were revealed.

3. Results

A hundred fifty-six total variants were found and 116 unique variants as shown in Tables 1 and 2. Among the 95 genomes we analyzed, 24 samples did not exhibit any variants except for missing starts and end base pairs. The distinct variants consist of 46 missense, 52 synonymous, 2 insertion, 1 deletion and 14 non-coding alleles in Fig. 1. The most common variants were 8782C > T(ORF1ab) in 13 samples, 28144T > C (ORF8) in 14 samples and 29095C > T (N) in 8samples. The occurrences of 8782C > T and 28144T > C coincide. 29095C > T is found in the subset of them. Both 8782C > T and 29095C > T are synonymous; however, 28144T > C causes amino acid to change L84S in ORF8. Notably, most of 8782C > T and 28144T > Cvariant substrains are found outside of Wuhan. For the 46 missense variants, 24 variants are found in ORF1ab, which is the longest ORF occupying 2/3 of the entire genome. ORF1ab is cleaved into many nonstructural proteins (NSP1-NSP16). Among NSP's, NSP3 has more variants in the analyzed samples. All noncoding mutations are located in 5′ UTR or 3′ UTR regions. In terms of base changes, the most frequently observed one is C > T as shown in Tables 1 and 2.

4. Discussion

The genetic information of any life is protected in its genome, and annotation is the initial step to interpret the sequence. The length of the SARS-CoV genome is over 30 Kb, while just a few coding genes appear not to accord with the general properties for the viral genome and the minimum grouping of hereditary data. In addition to these, it may have some non-structural proteins but lacking data at one place is needed. The absence probably results from their short existing-time before

Table 1
Coding mutation list detected in SARS-CoV-2 genomes.

Accession	Location-date	Nucleotide variation	Gene	Amino acid change	Mutation type
MT240479	04-03-2020/Pakistan Gilgit	1 1497G > A	Orf1ab		Synonymous mutation
MN996527	30/Dec/2019-China Wuhan	21316G > A	Orf1ab	D7018N	Missense
MN996527	30/Dec/2019-China Wuhan	24292A > G	S		Synonymous mutation
LC528232	10/Feb/2020-Japan	11083T > G	Orf1ab	L3606F	Missense
LC528232	10/Feb/2020-Japan	29642C > T	ORF10		Synonymous mutation
LR757995	05/Jan/2020-China Wuhan	28144T > C	ORF8	L84S	Missense
LR757998	12/26/2019-China Wuhan	6968C > A	Orf1ab	L2235I	Missense
LR757998	12/26/2019-China Wuhan	11749T > A	Orf1ab		Synonymous mutation
MN938384	1/10/2020-China Shenzhen	8782C > T	Orf1ab		Synonymous mutation
MN938384	1/10/2020-China Shenzhen	28144T > C	ORF8	L84S	Missense
MN938384	1/10/2020-China Shenzhen	29095C > T	N		Synonymous mutation
MN975262	11/Jan/2020-China	8782C > T	Orf1ab		Synonymous mutation
MN975262	11/Jan/2020-China	9534C > T	Orf1ab	T3090I	Missense
MN975262	11/Jan/2020-China	29095C > T	N		Synonymous mutation
MN975262	11/Jan/2020-China	28144T > C	ORF8	L84S	Missense
MN975262	11/Jan/2020-China	8782C > T	Orf1ab		Synonymous mutation
MN985325	19/Jan/2020-USA WA	28144T > C	ORF8	L84S	Missense
MN994467	23/Jan/2020-USA CA	1548G > A	Orf1ab	S428N	Missense
MN994467	23/Jan/2020-USA CA	8782C > T	Orf1ab		Synonymous mutation
MN994467	23/Jan/2020-USA CA	26729T > C	M		Synonymous mutation
MN994467	23/Jan/2020-USA CA	28077G > C	ORF8	V62L	Missense
MN994467	23/Jan/2020-USA CA	28144T > C	ORF8	L84S	Missense
MN994467	23/Jan/2020-USA CA	28792A > C	N		Synonymous mutation
MN994467	23/Jan/2020-USA CA	1912C > T	Orf1ab		Synonymous mutation
GWHABKF00000001	23/Dec/2019-China Wuhan	3778A > G	Orf1ab		Synonymous mutation
GWHABKF00000001	23/Dec/2019-China Wuhan	8388A > G	Orf1ab	N2708S	Missense
GWHABKF00000001	23/Dec/2019-China Wuhan	8987T > A	Orf1ab	F2908I	Missense
GWHABKK00000001	30/Dec/2019-China Wuhan	24325A > G	S		Synonymous mutation
GWHABKK00000001	30/Dec/2019-China Wuhan	21316G > A	Orf1ab	D7018N	Missense
GWHABKH00000001	30/Dec/2019-China Wuhan	6996T > C	Orf1ab	I2244T	Missense
GWHABKJ00000001	01/Jan/2019-China Wuhan	7866G > T	Orf1ab	G2534V	Missense
GWHABKM00000001	30/Dec/2019-China Wuhan	21137A > G	Orf1ab	K6958R	Missense
GWHABKM00000001	30/Dec/2019-China Wuhan	7016G > A	Orf1ab	G2251S	Missense
GWHABKO00000001	30/Dec/2019-China Wuhan	8001A > C	Orf1ab	D2579A	Missense
GWHABKO00000001	30/Dec/2019-China Wuhan	9534C > T	Orf1ab	T3090I	Missense
MT188341	05/Mar/2020-USA MN	6035A > G	Orf1ab		Synonymous mutation
MT188341	05/Mar/2020-USA MN	8782C > T	Orf1ab		Synonymous mutation
MT188341	05/Mar/2020-USA MN	16467A > G	Orf1ab		Synonymous mutation
MT188341	05/Mar/2020-USA MN	18060C > T	Orf1ab		Synonymous mutation
MT188341	05/Mar/2020-USA MN	21386insT	Orf1ab		Insertion

(*continued on next page*)

Table 1 (*continued*)

Accession	Location-date	Nucleotide variation	Gene	Amino acid change	Mutation type
MT188341	05/Mar/2020-USA MN	21388-21390insTT	Orf1ab		Insertion
MT188341	05/Mar/2020-USA MN	23185C > T	S		Synonymous mutation
MT188341	05/Mar/2020-USA MN	28144T > C	ORF8	L84S	Missense
MT188339	09/Mar/2020-USA MN	8782C > T	Orf1ab		Synonymous mutation
MT188339	09/Mar/2020-USA MN	17423A > G	Orf1ab	Y5720C	Missense
MT188339	09/Mar/2020-USA MN	18060C > T	Orf1ab		Synonymous mutation
MT188339	09/Mar/2020-USA MN	21386C > T	Orf1ab		Synonymous mutation
MT188339	09/Mar/2020-USA MN	22432C > T	S		Synonymous mutation
MT188339	09/Mar/2020-USA MN	28144T > C	ORF8	L84S	Missense
MT121215	02/Feb/2020-China Shanghai	6031C > T	Orf1ab		Synonymous mutation
MT123290	05/Feb/2020-China Guangzhou	15597T > C	Orf1ab		Synonymous mutation
MT123290	05/Feb/2020-China Guangzhou	29095C > T	N		Synonymous mutation
MT126808	2/28/2020-Brazil	26144G > T	ORF3a	G251V	Missense
MT066175	31/Jan/2020-Taiwan	8782C > T	Orf1ab		Synonymous mutation
MT066175	31/Jan/2020-Taiwan	28144T > C	ORF8	L84S	Missense
MT093571	07/Feb/2020-Sweden	13225C > G	Orf1ab		Synonymous mutation
MT093571	07/Feb/2020-Sweden	13226T > C	Orf1ab		Synonymous mutation
MT093571	07/Feb/2020-Sweden	17423A > G	Orf1ab	Y5720C	Missense
MT093571	07/Feb/2020-Sweden	23952T > G	S		Synonymous mutation
MT066156	30/Jan/2020-Italy	11083T > G	Orf1ab	L3606F	Missense
MT066156	30/Jan/2020-Italy	26144G > T	ORF3a	G251V	Missense
LC522975	20/JAN/2020-JAPAN	8782C > T	Orf1ab		Synonymous mutation
LC522975	20/JAN/2020-JAPAN	29095C > T	N		Synonymous mutation
LC522975	20/JAN/2020-JAPAN	28144T > C	ORF8	L84S	Missense
LC522975	20/JAN/2020-JAPAN	2662C > T	ORF1ab		Synonymous mutation
LC522974	20/JAN/2020-JAPAN	8782C > T	ORF1ab		Synonymous mutation
LC522974	20/JAN/2020-JAPAN	29095C > T	N		Synonymous mutation
LC522974	20/JAN/2020-JAPAN	28144T > C	ORF8	L84S	Missense
LC522974	20/JAN/2020-JAPAN	2662C > T	ORF1ab		Synonymous mutation
LC522973	20/JAN/2020-JAPAN	8782C > T	ORF1ab		Synonymous mutation
LC522973	20/JAN/2020-JAPAN	29095C > T	N		Synonymous mutation
LC522973	20/JAN/2020-JAPAN	3792C > T	ORF1ab	A1176V	Missense
LC522973	20/JAN/2020-JAPAN	29095C > T	N		Synonymous mutation
LC522973	20/JAN/2020-JAPAN	2662C > T	ORF1ab		Synonymous mutation
LC522973	20/JAN/2020-JAPAN	28144T > C	ORF8	L84S	Missense
LC522972	20/JAN/2020-JAPAN	29303C > T	N	P344S	Missense
LC522972	20/JAN/2020-JAPAN	25810C > G	ORF3a	L140V	Missense
LC522972	20/JAN/2020-JAPAN	11557G > T	ORF1ab	E3764D	Missense
LC522972	20/JAN/2020-JAPAN	15324C > T	ORF1ab		Synonymous mutation
LC521925	21/JAN/2020-JAPAN	1912C > T	ORF1ab		Synonymous mutation
LC521925	21/JAN/2020-JAPAN	18512C > T	ORF1ab	P6083L	Missense
LC521925	21/JAN/2020-JAPAN	359_382del	ORF1ab	G32_L39del	Deletion
MN988713	21/JAN/2020-USA Chicago	24034C > T	S		Synonymous mutation
MN988713	21/JAN/2020-USA Chicago	26729T > C	M		Synonymous mutation
MN988713	21/JAN/2020-USA Chicago	8782C > T	ORF1ab		Synonymous mutation
MN988713	21/JAN/2020-USA Chicago	490T > A	ORF1ab	D75E	Missense
MN988713	21/JAN/2020-USA Chicago	3177C > T	ORF1ab	P971L	Missense
MN988713	21/JAN/2020-USA Chicago	28854C > T	N	S194L	Missense
MN988713	21/JAN/2020-USA Chicago	28077G > C	ORF8	V62L	Missense
MN988713	21/JAN/2020-USA Chicago	28144T > C	ORF8	L84S	Missense
MN997409	21/JAN/2020-USA Arizona	8782C > T	ORF1ab		Synonymous mutation
MN997409	21/JAN/2020-USA Arizona	29095C > T	N		Synonymous mutation

(*continued on next page*)

Table 1 (*continued*)

Accession	Location-date	Nucleotide variation	Gene	Amino acid change	Mutation type
MN997409	21/JAN/2020-USA Arizona	11083G > T	ORF1ab	L3606F	Missense
MN997409	21/JAN/2020-USA Arizona	28144T > C	ORF8	L84S	Missense
MT072688	26/JAN/2020-USA: Massachussetts	24034C > T	S		Synonymous mutation
NMDC60013002-09	01/JAN/2019-China Wuhan	27493C > T	ORF7a	P34S	Missense
NMDC60013002-09	01/JAN/2019-China Wuhan	28253C > T	ORF8		Synonymous mutation
NMDC60013002-10	30/Dec/2019-China Wuhan	20679G > A	ORF1ab		Synonymous mutation
NMDC60013002-01	30/Dec/2019-China Wuhan	11764T > A	ORF1ab	N3833K	Missense
NMDC60013002-06	30/Dec/2019-China Wuhan	24325A > G	S		Synonymous mutation
NMDC60013002-04	05/Dec/2019-China Wuhan	28144T > C	ORF8	L84S	Missense

Table 2
Non-coding mutation list detected in SARS-CoV-2 genomes.

Accession	Location-date	Nucleotide variation	UTR type
MT240479	04-03-2020/Pakistan Gilgit	241C > T	5 UTR
MT123290	05/Feb/2020-China Guangzhou	4A > T	5 UTR
MT007544	25/Jan/2020-Australia Victoria	29749-29759del	3 UTR
NMDC60013002-07	07/JAN/2019-China Wuhan	29869del	3 UTR
NMDC60013002-04	05/Dec/2019-China Wuhan	29856T > A	3 UTR
NMDC60013002-04	05/Dec/2019-China Wuhan	29854C > T	3 UTR
NMDC60013002-04	05/Dec/2019-China Wuhan	16C > T	5 UTR
MT049951	17/Jan/2019-China Yunnan	75C > A	5 UTR
LC522975	20/JAN/2020-JAPAN	29705G > T	3 UTR
GWHABKG00000001	30/Dec/2019-China Wuhan	124G > A	5 UTR
GWHABKG00000001	30/Dec/2019-China Wuhan	120T > C	5 UTR
GWHABKG00000001	30/Dec/2019-China Wuhan	119C > G	5 UTR
GWHABKG00000001	30/Dec/2019-China Wuhan	112T > G	5 UTR
GWHABKG00000001	30/Dec/2019-China Wuhan	111T > C	5 UTR
GWHABKG00000001	30/Dec/2019-China Wuhan	104T > A	5 UTR

decomposition. In this study, we worked to find the extent of molecular variation between the recently sequenced genomes of SARS-CoV-2.

Numerous investigations have depicted that ORFs and ACE2 genes play a key role during novel coronavirus disease (Koyama et al., 2020; Kirchdoerfer and Ward, 2019; Van der Meer et al., 1998; Wan et al., 2020). So in our study, 156 total variants were found and 116 unique variants (Tables 1 and 2). Among the 95 genomes we analyzed, 24 samples did not exhibit any variants except for missing starts and end base pairs. Additionally, the distinct variants consist of 46 missense, 52 synonymous, 2 insertions, 1 deletion, 14 non-coding alleles (Tables 1 and 2). Most common variants were 8782C > T(ORF1ab) in 13 samples, 28144T > C (ORF8) in 14 samples and 29095C > T (N) in 8 samples.The occurrences of 8782C > T and 28144T > C coincide. 29095C > T is found in the subset of them. Both 8782C > T and 29095C > T are synonymous; however, 28144T > C causes amino acid to change L84S in ORF8. It is notable that most of 8782C > T and 28144T > C variant substrains are found outside of Wuhan. For the 46 missense variants, 24 variants are found in ORF1ab, which is the longest ORF occupying 2/3 of the entire genome. ORF1ab is cleaved into many nonstructural proteins (NSP1-NSP16). Among NSP's, NSP3 has more variants in the analyzed samples. All noncoding mutations are located in 3′UTR or 5′UTR. In terms of base changes, the most frequently observed one is C > T (Tables 1 and 2).

The replicase enzyme is displayed as two polyproteins (ORF1a and ORF1ab), which are prepared into 12 nonstructural proteins by three viral proteases (Van der Meer et al., 1998). This ORF1ab polyprotein includes the nsps 1–3 proteins. This area of ORF1ab is the most important factor among coronaviruses (Wan et al., 2020). Many researchers found the relationship between ORFs with COVID-19 i.e.8782C > T(ORF1ab) and 28144T > C (ORF8) are available among genomic databases (Kirchdoerfer and Ward, 2019; Koyama et al., 2020). Hence, it will be clinically significant to break down the biological function of the particular protein ORF1ab in SARS-CoV-2.

Orf8 protein of SARS-CoV-2 doesn't contain a known useful motif or region. A total motif VLVVL (amino corrosive 75–79) has been found in SARS-CoV orf8b which was appeared to trigger intracellular stress pathways and enact NOD-like receptor family pyrin region containing-3 (NLRP3) (Shi et al., 2019). Moreover, multiple arrangements with different coronavirus ORF8 sequences propose that L84 related to 28144T > C (L84S) isn't preserved (Koyama et al., 2020). Thusly, it will be critical to examine the biological function of the particular protein (orf8) in SARS-CoV-2.

ORF10 is a short protein or peptide of length 38 deposits. Koyama et al. depicted that COVID-19 is ORF10 which doesn't have any comparative proteins in the NCBI repository. This one of a kind protein can be used to distinguish the infection more rapidly than PCR based strategies (Koyama et al., 2020), but the further characterization of this protein is strongly required.

Another study demonstrated that NCBI had displayed new annotations for orf1ab as of late. NSP6 is the main contrast and it is considered as a putative protein (Koyama et al., 2020). So, they held the NSP annotations. They further referenced that 12 remarkable variations in NSP3 protein in ORF1ab. Thus concluded that there was a basic connection between the nsp3 association and the inception of coronavirus infection (Hurst et al., 2013a). Besides, they investigated that NSP3 contains the papain-like protease and is regarded as significant for SARS infection (Niemeyer et al., 2018). Variations found in subjects began from Wuhan are situated in either TM1 or Y space which is profoundly saved (Hurst et al., 2013a, 2013b).

Sawicki et al. performed sequencing of ORF1 from a huge available data that was established in labs (Sawicki et al., 2005). The report distinguished single point transformations coming from

nonsynonymous substitutions in nsps 4, 5, 10, 12, 14 and 16. The collective outcomes recommend that the ORF1b nsp 12, 14 and 16 proteins characterize particular cistrons, while the distorted ORF1a proteins nsps 4, 5, and 10 form a compartment together at the location of the coding sequence of ORF1ab (Graham et al., 2008). Notably, of the eight announced mutations in MHV, seven of the influenced amino acid deposits are correlated with SARS-CoV (Graham et al., 2008). This methodology may permit the determination of phenotypic travelers that will distinguish by protein interactions. Also, it will permit the progressions to be acquainted in SARS-CoV with deciding whether the ts phenotype can be reproduced in that foundation, with the chance of quickly building up a board of SARS-CoV. Curiously, not only is the slow-growth branch dominated by travelers, but the COVID-19 lineages appear to be phylogenetically related to each other, suggesting an exposure point for these individuals that are distinct from the rest of the population.

5. Conclusion

The fast increment of cases is giving more genomes that may give some visibility and proof of populace structure, especially of the chance of various presentations of COVID-19 into the human population. A comprehension of the biological reservoirs conveying these infections, and how the course to introduce has been carrying them into contact with human beings will be critical to comprehend future risks for novel diseases. This study showed how the disease spread among the travelers. This fight against COVID-19 will be a long one until we develop vaccines or effective treatments. However, we believe that collecting and sharing knowledge on variants will be effective. We should continue to be vigilant for the emergence of new variants or substrains and data should be gathered at one place for better understanding.

CRediT authorship contribution statement

Rozhgar A. Khailany:Conceptualization, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing.**Muhamad Safdar**:Conceptualization, Visualization, Writing - original draft, Writing - review & editing.**Mehmet Ozaslan**:Conceptualization, Supervision, Visualization, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

We highly appreciate many members of the frontline medical and nursing staff who demonstrated selfless and heroic devotion to duty in the face of this outbreak.

References

Adhikari, S., Meng, S., Wu, Y., et al., 2020. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. Infect Dis Poverty 9, 29.
Arvestad, L., 2018. alv: a console-based viewer for molecular sequence alignments. Journal of Open Source Software 3 (31), 955.
Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395, 507–513.
De Wit, E., van Doremalen, N., Falzarano, D., Munster, V.J., 2016. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 14, 523–534.
Ge, X.Y., et al., 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503, 535–538.
Graham, R.L., Sparks, J.S., Eckerle, L.D., Sims, A.C., Denison, M.R., 2008. SARScoronavirus replicase proteins in pathogenesis. Virus Res. 133 (88–10).
Hurst, K.R., Koetzner, C.A., Masters, P.S., 2013a. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. J. Virol. 87 (16), 9159–9172.
Hurst, Kelley R., Koetzner, Cheri A., Paul, S., 2013b. J. Virol. 87 (16), 9159. https://doi.org/10.1128/JVI.01275-13.
Kirchdoerfer, R.N., Ward, A.B., 2019. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat. Commun. 10, 2342.
Koyama, T., Platt, D., Parida, L., 2020. Variant analysis of COVID-19 genomes. Bull. World Health Organ. https://doi.org/10.2471/BLT.20.253591.
Li, F., Li, W., Farzan, M., Harrison, S.C., 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309, 1864–1868.
Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med. https://doi.org/10.1056/NEJMoa2001316.
Lu, R., Zhao, X., Li, J., et al., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395, 565–574.
Niemeyer, D., Mosbauer, K., Klein, E.M., Sieberg, A., Mettelman, R.C., Mielech, A.M., et al., 2018. The papain-like protease determines a virulence trait that varies among members of the SARS-coronavirus species. PLoS Pathog. 14 (9), e1007296.
Oostra, M., de Haan, C.A., Rottier, P.J., 2007. The 29-nucleotide deletion present in human but not in animal severe acute respiratory syndrome coronaviruses disrupts the functional expression of open reading frame 8. J. Virol. 81, 13876–13888.
Ozaslan, M., Safdar, M., Kilic, I.H., Khailany, R.A., 2020. Practical measures to prevent COVID-19: a mini-review. J. Biol. Sci. 20 XX-XX.
Payne, D.C., et al., 2018. Multihospital outbreak of a Middle East respiratory syndrome coronavirus deletion variant, Jordan: a molecular, serologic, and epidemiologic investigation. Open Forum Infect Di 5.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16 (6), 276–277.
Sawicki, S.G., Sawicki, D.L., Younker, D., Meyer, Y., Thiel, V., Stokes, H., Siddell, S.G., 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. PLoSPathog 1 (4), e39.
Shi, C.S., Nabar, N.R., Huang, N.N., Kehrl, J.H., 2019. SARS-Coronavirus Open Reading Frame-8b triggers intracellular stress pathways and activates NLRP3 inflammasomes. Cell Death Discov 5, 101.
Van der Meer, Y., van Tol, H., Locker, J.K., Snijder, E.J., 1998. ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. J. Virol. 72, 6689–6698.
Wan, Y., Shang, J., Graham, R., Baric, R.S., Li, F., 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. J. Virol. 94https://doi.org/10.1128/JVI.00127-20. (e00127-20).
Yang, X.L., et al., 2015. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. J. Virol. 90, 3253–3256.