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.
Genomic characterization of a novel SARS-CoV-2

Rozhgar A. Khailany\textsuperscript{a,⁎}, Muhamad Safdar\textsuperscript{b}, Mehmet Ozaslan\textsuperscript{c}

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

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

| S Gene | ORF1ab Polyprotein | ENv | ORF3a Protein | ORF4 Protein | ORF5a Protein | ORF5b Protein | ORF7a Protein | ORF7b Protein | ORF8 Protein |
|--------|---------------------|-----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Surface Glycoprotein (S) | Membrane Envelope (E) | Nucleocapsid (NC) | ORF3a ORF6 ORF7a ORF7b ORF8 | Nucleocapsid Protein (N) | ORF1ab | ORF3a | ORF4 | ORF5a | ORF5b | ORF7a | ORF7b | ORF8 |
| 20knt | 38247 | 16knt | 156 | 189 | 220 | 200 | 192 | 182 | 172 | 162 | 152 | 142 | 132 |

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

SARS-CoV-2 Complete Genome (29903 Nucleotides)
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 | 1497G > A            | Orf1ab | D7018N           | Missense      |
| MN996527  | 30/Dec/2019-China Wuhan | 21316G > A           | Orf1ab |                  |               |
| MN966527  | 30/Dec/2019-China Wuhan | 24292A > G           | S     |                  | Synonymous mutation |
| LC528232  | 10/Feb/2020-Japan      | 11083T > G           | Orf1ab | L3606F           | Missense      |
| LR757995  | 05/Jan/2020-China Wuhan | 29642C > T           | ORF10 |                  | Synonymous mutation |
| 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           | ORF8  | L84S             | Missense      |
| MN938384  | 1/10/2020-China Shenzhen | 28144T > C           | ORF8  | L84S             | Missense      |
| MN975262  | 11/Jan/2020-China Shenzhen | 8782C > T           | Orf1ab |                  | Synonymous mutation |
| MN975262  | 11/Jan/2020-China Shenzhen | 9534C > T           | Orf1ab | T3090I           | Missense      |
| MN975262  | 11/Jan/2020-China Shenzhen | 29095C > T           | N     |                  | Synonymous mutation |
| MN975262  | 11/Jan/2020-China Shenzhen | 8782C > T           | ORF8  | L84S             | Missense      |
| 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      |
| GWHABKJK00000001 | 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)
| 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 |
| MT188339  | 09/Mar/2020-USA  | MN 28144T > C | ORF8 L84S | Missense |
| MT188339  | 09/Mar/2020-USA  | MN 18060C > T | S | 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 |
| 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 |
| MT093571  | 07/Feb/2020-Sweden | 23952T > G | S | Synonymous mutation |
| MT066156  | 30/Jan/2020-Italy | 11083T > G | Orf1ab | L3606F |
| MT066156  | 30/Jan/2020-Italy | 26144G > T | ORF3a | G251V |
| 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 |
| LC522975  | 20/JAN/2020-JAPAN | 28144T > C | Orf1ab | Synonymous mutation |
| LC522975  | 20/JAN/2020-JAPAN | 23952T > T | N | Synonymous mutation |
| LC522975  | 20/JAN/2020-JAPAN | 2662C > T | Orf1ab | Synonymous mutation |
| LC522975  | 20/JAN/2020-JAPAN | 28077G > C | ORF8 V62L | Missense |
| LC522975  | 20/JAN/2020-JAPAN | 28144T > C | ORF8 L84S | Missense |
| LC522975  | 20/JAN/2020-JAPAN | 2662C > T | ORF1ab | Synonymous mutation |
| LC522975  | 20/JAN/2020-JAPAN | 28144T > C | Orf1ab | Synonymous mutation |
| LC522975  | 20/JAN/2020-JAPAN | 28144T > C | Orf1ab | G32139del |
| MN988713  | 21/JAN/2020-USA Chicago | 24034C > T | S | Deletion |
| MN988713  | 21/JAN/2020-USA Chicago | 28729T > C | M | Synonymous mutation |
| MN988713  | 21/JAN/2020-USA Chicago | 8782C > T | Orf1ab | Synonymous mutation |
| MN988713  | 21/JAN/2020-USA Chicago | 4907 > A | Orf1ab | D75E |
| MN988713  | 21/JAN/2020-USA Chicago | 3177C > T | Orf1ab | P971L |
| MN988713  | 21/JAN/2020-USA Chicago | 28854C > T | N | S194L |
| MN988713  | 21/JAN/2020-USA Chicago | 28077G > C | Orf1ab | V62L |
| MN988713  | 21/JAN/2020-USA Arizona | 28144T > C | Orf1ab | L84S |
| MN997409  | 21/JAN/2020-USA Arizona | 28144T > C | Orf1ab | Synonymous mutation |
| MN997409  | 21/JAN/2020-USA Arizona | 29095C > T | N | Synonymous mutation |

(continued on next page)
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 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 NSPs, 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 (sars-cov-2). For instance, 28144T > C (ORF8) is preserved (Koyama et al., 2020). Thusly, it will be clinically significant to break down the biological function of the particular protein ORF1ab in SARS-CoV-2.

The noncoding mutation list detected in SARS-CoV-2 genomes. (continued)

| Accession | Location-date | Nucleotide variation | Gene | Amino acid change | Mutation type |
|-----------|---------------|----------------------|------|------------------|---------------|
| MN997409  | 21/JAN/2020-USA | 11093G > T | ORF1ab | L3606F | Missense |
| MT072688  | 26/JAN/2020-USA | 28144T > C | ORF1ab | L84S | Missense |
| NMD0013002-09 | 01/JAN/2019-China | 28144T > C | ORF1ab | L84S | Missense |
| NMD0013002-09 | 01/JAN/2019-China | 28144T > C | ORF1ab | L84S | Missense |
| NMD0013002-10 | 30/Dec/2019-China | 28144T > C | ORF1ab | L84S | Missense |
| NMD0013002-01 | 30/Dec/2019-China | 28144T > C | ORF1ab | L84S | Missense |

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
The fast increment of cases is giving more genomes that may give some visibility and proof of populace structure, especially of the chance for 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 vigilante 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, 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. SARS coronavirus 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.

Nieh, N., 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.

Onostra, M., de Haan, C.A., Rottier, P.J., 2007. The 29-nucleotide deletion present in the ORF1ab of the 2003 severe acute respiratory syndrome coronavirus: implications for virus origins and receptor binding. J. Virol. 81, 13876–13888.

Ozçan, 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 Dis 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. PLoS Pathog 1 (4), e39.

Shi, C.S., Nabar, N.R., Huang, N.N., Kehrl, J.H., 2020. SARS-CoV-2 spike deletion mutant influences amino acid deposits are correlated with SARS-CoV (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).

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.