



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.



## Genetic characterization of structural and open reading Fram-8 proteins of SARS-CoV-2 isolates from different countries

Abdullah Shah<sup>a</sup>, Farooq Rashid<sup>b</sup>, Abdul Aziz<sup>c</sup>, Amin Ullah Jan<sup>a</sup>, Muhammad Suleman<sup>d,\*</sup>

<sup>a</sup> Department of Biotechnology, Shaheed Benazir Bhutto University Sheringal, Dir, Pakistan

<sup>b</sup> School of Public Health, Department of Epidemiology, Southern Medical University, Dermatology Hospital, Guangzhou, China

<sup>c</sup> Molecular Biology Research Center, School of Life Sciences, Central South University, Changsha, China

<sup>d</sup> Centre for Biotechnology and Microbiology, University of Swat, Pakistan

### ARTICLE INFO

#### Keywords:

SARS-CoV-2  
COVID-2  
Mutation  
Phosphorylation  
Serine/Threonine kinases

### ABSTRACT

Since December 2019, a severe pandemic of pneumonia, COVID-19 associated with a novel coronavirus (SARS-CoV-2), have emerged in Wuhan, China and spreading throughout the world. As RNA viruses have a high mutation rate therefore we wanted to identify whether this virus is also prone to mutations. For this reason we selected four major structural (Spike protein (S), Envelope protein (E), Membrane glycoprotein (M), Nucleocapsid phosphoprotein (N)) and ORF8 protein of 100 different SARS-CoV-2 isolates of fifteen countries from NCBI database and compared these to the reference sequence, Wuhan NC\_045512.2, which was the first isolate of SARS-CoV-2 that was sequenced. By multiple sequence alignment of amino acids, we observed substitutions and deletion in S protein at 13 different sites in the isolates of five countries (China, USA, Finland, India and Australia) as compared to the reference sequence. Similarly, alignment of N protein revealed substitutions at three different sites in isolates of China, Spain and Japan. M protein exhibits substitution only in one isolates from USA, however, no mutation was observed in E protein of any isolate. Interestingly, in ORF8 substitution of Leucine, a nonpolar to Serine a polar amino acid at same position (aa84 L to S) in 23 isolates of five countries i.e. China, USA, Spain, Taiwan and India were observed, which may affect the conformation of peptides. Thus, we observed several mutations in the isolates thereafter the first sequencing of SARS-CoV-2 isolate, NC\_045512.2, which suggested that this virus might be a threat to the whole world and therefore further studies are needed to characterize how these mutations in different proteins affect the functionality and pathogenesis of SARS-CoV-2.

### 1. Introduction

An outbreak that started in Early December 2019 in Wuhan City, Hubei Province, China has been claiming the lives of many people there and, thereafter, has created an alarming situation around the globe. The agent causing this pandemic of pneumonia is the novel Betacoronavirus named by the World Health Organization (WHO) as “Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2)” and the disease caused by the virus has been named as COVID 19 (Zhu et al., 2020).

Since its first appearance in Wuhan, China, 82,960 individuals have been reported positive for the coronavirus (SARS-CoV-2) and the death toll recorded in China has reached 4634 as of May 19, 2020. The pace of spreading of the outbreak merits attention and as of March 22, 2020,

the virus has spread to more than 200 countries causing over 4,911,700 confirmed positive cases and engulfing 320,454 individuals around the world (WHO, 2020).

Host to host transmission of the SARS-CoV-2 is extremely high and the symptoms manifested by most patients with COVID-19 include fever, coughing, dyspnea, myalgia, shortness of breath and radiological evidence of ground-glass lung opacities compatible with atypical pneumonia (Huang et al., 2020; Lu et al., 2020a). However, some patients with asymptomatic or mildly symptomatic cases have also been reported. The recorded mortality rate of the infection of SARS-CoV-2 is about 3% (Huang et al., 2020; Lu et al., 2020a; Chen et al., 2020).

The coronaviruses, responsible for playing the havoc around the world, are enveloped single stranded RNA viruses with a wide host

**Abbreviations:** Name, Abbreviations; Severe Acute Respiratory Syndrome-Coronavirus-2, SARS-CoV-2; coronavirus disease of 2019, COVID-19; angiotensin converting enzyme 2, ACE2; surface glycoprotein, S; envelope protein, E; membrane glycoprotein, M; nucleocapsid phosphoprotein, N; nonstructural proteins, NSPs; World Health Organization, WHO; open reading frame, ORF

\* Corresponding author at: Centre for Biotechnology and Microbiology, University of Swat, Pakistan

E-mail address: [suleman@uswat.edu.pk](mailto:suleman@uswat.edu.pk) (M. Suleman).

<https://doi.org/10.1016/j.genrep.2020.100886>

Received 20 May 2020; Received in revised form 4 July 2020; Accepted 11 September 2020

Available online 14 September 2020

2452-0144/ © 2020 Elsevier Inc. All rights reserved.

spectrum including humans, birds and mammals. These viruses can cause respiratory, enteric, hepatic and neurological diseases and gastrointestinal tract infections. Genetically, the viruses have been classified into four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus (Chen et al., 2020; Guan et al., 2020; Weiss and Leibowitz, 2011). The first two genera, Alphacoronavirus and Betacoronavirus, are chiefly associated with infecting mammals while the later two, Gammacoronavirus and Deltacoronavirus, are mostly found in infections associated with birds (Chan et al., 2015; Li, 2016).

Humans can be infected by six different kinds of coronavirus species and can cause disease in them. These include HCoV-NL63 and HCoV229E belonging to the Alphacoronavirus genus while HCoV-OC43 and HCoVHKU1 to the Betacoronavirus genus. These viruses are the most prevalent causing infection in humans producing symptoms of common cold in immunocompetent individuals. The other two strains, which have been found associated with fatal illness, are zoonotic in origin and include Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) belonging to Betacoronavirus genus (Tang et al., 2015).

Coronaviruses for the first time came under spotlight globally when a SARS, Severe Acute Respiratory Syndrome, pandemic outbreak in 2002 and 2003 in Guangdong Province, China. Another episode of the outbreak was witnessed in 2012 in the form of MERS, a severe respiratory disease outbreak in the Middle East and currently since December 2019 in China, the COVID-19 eruption has taken the world by storm (Song et al., 2019; China CDC, 2020). It is believed that SARS-CoV and MERS-CoV might have transmitted from bats to palm civets or dromedary camels and finally to humans and both are considered as highly pathogenic (Cui et al., 2019; Guan et al., 2003; Drosten et al., 2014).

The genome of the SARS-CoV-2 has been sequenced. The genome has been described to have ~29.8 kb and contained 14 ORFs which encode 27 proteins. The two genes, ORF1ab and ORF1a, encoding two long polypeptides, pp1ab and pp1a respectively, are located on the 5'-terminus of the genome. Proteolytically, these two polypeptides, pp1ab and pp1a, have further been processed into 14 nonstructural proteins (nsps) including nsp1 to nsp10 and nsp12 to nsp16. The four structural proteins constituting the main envelop of the virus and eight accessory proteins are encoded by the genes located on the 3'-terminus. The Spike surface glycoprotein (S), Small envelope protein (E), Membrane protein (M) and Nucleocapsid protein (N) are constituted by the four structural proteins. The eight accessory proteins include 3a, 3b, p6, 7a, 7b, 8b, 9b and ORF14 (Wu et al., 2020a).

The binding of the virus to the receptors on the host cell and fusion with the cell membrane is aided by the Spike surface glycoprotein (Li, 2016; Kandeel et al., 2018), while the N protein is assigned to interact with the viral RNA to form the ribonucleoproteins. The E proteins in collaboration with the M protein helps in viron assembly. The E protein is also known to function in comprising ion channel actions (Kandeel et al., 2018; Wu et al., 2020b; Chan et al., 2020; Lu et al., 2020b).

Recently, the SARS-CoV-2 outburst has raised the eyebrows of all and sundry around the world since the virus undergoes a potent immunologic pressure in humans and, consequently, may acquire the capability to outwit the human's immune system by accumulating mutations. Such mutations, with the passage of time, could possibly bring fluctuation in viral infectivity, virulence level and transmissibility in terms of pace and host selection (Lucas et al., 2001). Hence, it is indispensable to undergo a circumspective study to analyze the pattern and frequency of mutations thus far accumulated by the virus in order to project its future implications. Fortunately, the entire genome of the virus has been sequenced by the worldwide scientific community which could be utilized to carry out the current study. Taking the advantage of genomic information currently available, the genome of SARS-CoV-2 isolates from fourteen different countries (China, USA, Spain, Japan, Pakistan, India, Nepal, Australia, Italy, Brazil, Sweden, Finland, Viet

Nam and Taiwan) have been analyzed for mutation in four major structural proteins (Spike glycoprotein (S), Envelope protein (E), Membrane glycoprotein (M), Nucleocapsid phosphoprotein (N) and an accessory protein, ORF8, which are compared to reference sequence, Wuhan NC\_045512.2 isolate.

## 2. Materials and methods

### 2.1. Gene data collection

The amino acids sequences of four structural protein (Spike glycoprotein (S), Nucleocapsid (N), Envelop (E), membrane (M)) and ORF8 gene of different isolates of SARS-CoV-2 from fifteen different countries (China, USA, Spain, Japan, Pakistan, India, Nepal, Australia, Italy, turkey, Brazil, Sweden, Finland, Viet Nam and Taiwan) were downloaded from the NCBI (<https://www.ncbi.nlm.nih.gov/>).

### 2.2. Multiple sequence alignment

The DNASTAR. Lasergene (7.1) and Clustal W program of the MEGA software (7.1) were used to conduct multiple sequence alignment of four structural (S, E, M, and N) and ORF8 proteins of 100 targeted SARS-CoV-2 isolates from different countries (China, USA, Spain, Japan, Pakistan, India, Nepal, Australia, Italy, turkey, Brazil, Sweden, Finland, Viet Nam and Taiwan). For mutational analysis the previously reported isolates of SARS-CoV-2 (Wuhan NC\_045512.2) was used as reference sequence (Ceraolo and Georgi, 2020).

## 3. Results

### 3.1. Mutational analyses of the SARS-CoV-2 structural proteins

The present study carried out the amino acids sequence alignment of four structural proteins (Spike surface glycoprotein (S), Membrane glycoprotein (M), Envelope protein (E), Nucleocapsid phosphoprotein (N)) and one ORF8 protein of the novel SARS-CoV-2.

### 3.2. Mutational analysis identified substitution and deletion mutations in Spike surface glycoprotein (S)

The Spike surface glycoprotein of SARS-CoV-2 is 1273 amino acid in length and play an essential role in binding of the virus to the receptors, angiotensin converting enzyme 2 (ACE2) on the host cell and initiate the infection (Tang et al., 2015; Wu et al., 2020b). The present study identified amino acids substitutions and deletion in some isolates of five countries as compared to the reference sequence (NC\_045512.2). Two isolates of China with accessions numbers MT 049951 and MT 226610 have substitution at (aa28 Y to N) and (aa74 N to K) positions respectively. Similarly, the amino acids substitution were observed in ten isolates of USA, with accessions numbers, MT 159716 (aa157 F to L), MT 246490 (aa614 D to G), MT 251973 (aa 614 D to G), MT 251974 (aa614 D to G), MT 251976 (aa614 D to G), MT 251979 (aa 614 D to G), MT 163720 (aa655 H to Y), MT 093571 (aa797 F to C), MT 050493 (aa930 A to V), while two substitutions were observed at position (aa5 L to F) and (aa 476 G to S) in isolate having accession MT 246488. (Fig. 1). Likewise, the isolate of Finland (MT 020781) and Australia (MT 007544) have substitution at position (aa49 H to Y) and (aa 247 S to R) respectively. However isolate from India (MT 012098) has deletion at position (aa 145 Y) as well as a substitution at (aa 408 R to I) (Fig. 1).

### 3.3. Mutational analysis identified the substitution mutations in Nucleocapsid phosphoprotein (N)

Mutational analysis of the N protein (419 amino acid residues) from fifteen countries (100 isolates) revealed the amino acids substitutions at

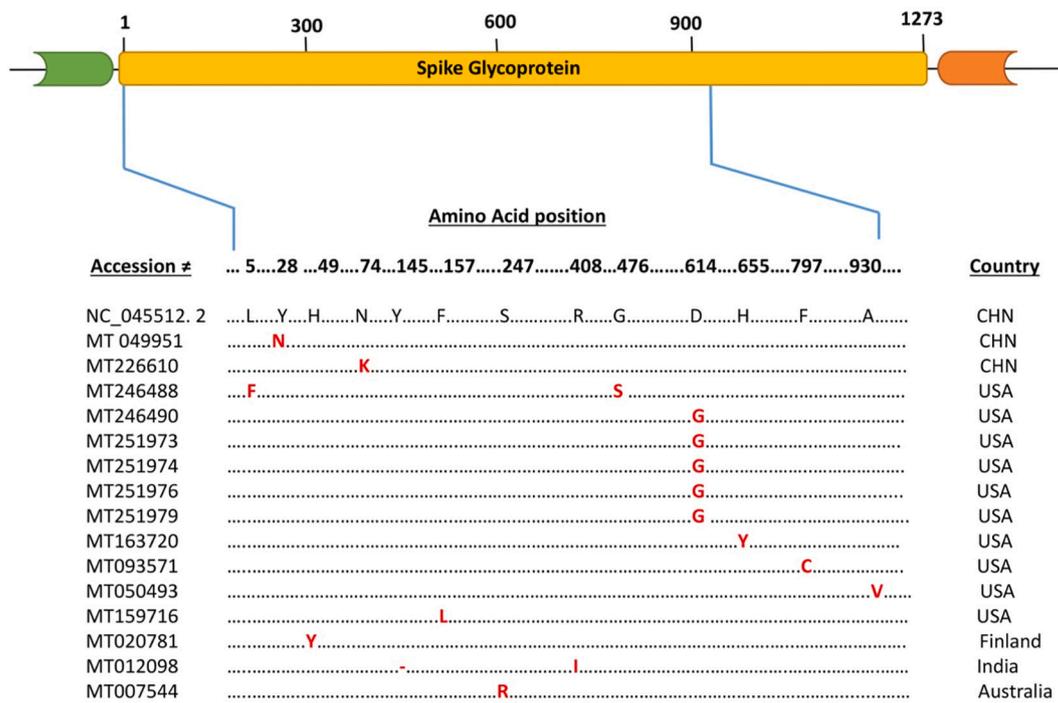


Fig. 1. Schematic diagram and Multiple Sequence Alignment of the structural, Spike surface glycoprotein (S) of SARS-CoV-2. Dot represents identical amino acid, dash represents deletion and the letters highlighted in red represents substitution as compared to the reference sequence (NC\_045512.2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

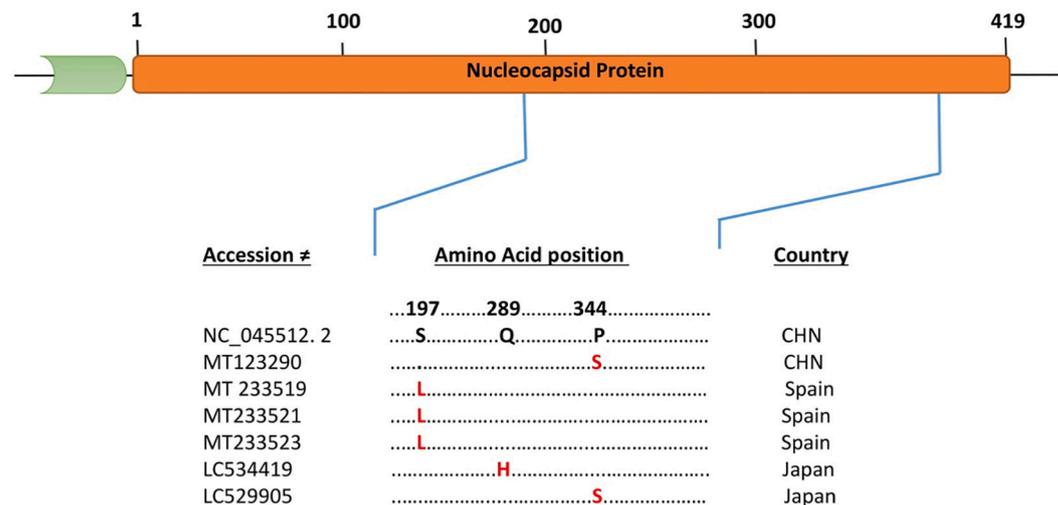


Fig. 2. Schematic diagram and Multiple Sequence Alignment of the structural, Nucleocapsid protein (N) of SARS-CoV-2. Dot represent identical amino acid and the letters highlighted in red represents substitution as compared to the reference sequence (NC\_045512.2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

three different sites in six isolates (Fig. 2). Interestingly, three isolates of Spain with accessions numbers MT 233519, MT 233521 and MT 233523 have substitution at the same position (aa 197 S to L), while two isolates one from Japan (LC429905) and second from China (MT 123290) have substitution at the same location (aa 344 P to S). Similarly, one isolate from Japan (LC534419) has substitution at (aa289 Q to H) (Fig. 2).

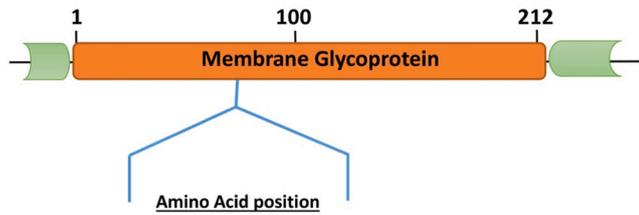
### 3.4. Mutational analysis of Envelope protein (E)

The Envelope protein of SARS-CoV-2 consist of 75 amino acids residues and helps in virions assembly (Lucas et al., 2001; Zhu et al., 2018). The alignment of the amino acids sequence of the E protein of all

hundred isolates, with corresponding E protein of the reference isolates (NC\_045512.2) revealed that the SARS-COV-2, E protein is relatively conserved where no mutation was found figure not shown.

### 3.5. Substitution mutation was identified in Membrane glycoprotein (M) in one isolate of USA

Membrane glycoprotein of SARS-CoV-2 consists of 222 amino acids residues and involved in the assembly of new virus particles (Lu et al., 2020b; Hu et al., 2018). The present study revealed the substitution of amino acid only in one isolate from USA, with accession number MT 163721 at position (aa70 V to I) as compared to the reference counterpart (NC\_045512.2) (Fig. 3).



Accession #	70	Country
NC_045512.2	V	CHN
MT 163721	I	USA

Fig. 3. Schematic diagram and Multiple Sequence Alignment of the structural, Membrane protein (M) of SARS-CoV-2. Dot represent identical amino acid and the letters highlighted in red represents substitution as compared to the reference sequence (NC\_045512.2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 3.6. Characterization of ORF8 of SARS-CoV-2 protein identified substitution of Leucine to Serine

The ORF8 gene encode 121 amino acid residues (Wu et al., 2020b). Interestingly, in ORF8 we found substitution of Leucine, a nonpolar to Serine a polar amino acid at the same position (aa84 L to S) in 15 USA isolates (MT 163717, MT 163718, MT 163719, MT 163720, MT 163721, MT 246472, MT 246473, MT 246474, MT 246488, MT 246489, MT 251975, MT 251977, MT 251978, MT 251980), while one

isolate MT 251972 have two substitutions (aa84 L to S and aa112 H to Q). Three Spanish isolates (MT 233519, MT 233521, MT 233523), three Chinese isolates (MT 049951, MT 226610, MT 123292) and one from each Taiwan (MT 066175) and India (MT 050493) have the same substitutions at position (aa84 L to S) as compared to the reference counterpart (NC\_045512.2). This substitution of nonpolar amino acid by a polar one can affect the conformation of the peptide (Fig. 4).

### 4. Discussion

SARS-CoV-2 are enveloped single stranded RNA virus having genome of about 29.8 kb and was annotated to possess 14 ORFs, encoding 27 proteins (Wu et al., 2020b). The four major structure proteins are the Spike surface glycoprotein (S), Envelope protein (E), Membrane glycoprotein (M), and Nucleocapsid phosphoprotein (N). The Spike surface glycoprotein plays an essential role in binding of virus to angiotensin converting enzyme 2 (ACE2) receptors on the host cell, and thus crucial for determining host tropism and transmissibility (Li, 2016; Zhu et al., 2018). The N protein is one of the predominantly expressed proteins during the early stages of infection and interacts with viral RNA to form the ribonucleoprotein. This protein has been an attractive diagnostic tool due to the initiation of strong immune response against it. The E protein helps in virions assembly, while M protein shares in the assembly of new virus particles (Wu et al., 2020b; Chan et al., 2020; Lu et al., 2020b).

RNA viruses have a high mutation rate and thus, are prone to evolve resistance to drugs and escape from immune surveillance (Su et al., 2016). Thus a constant surveillance of mutations arising is needed.

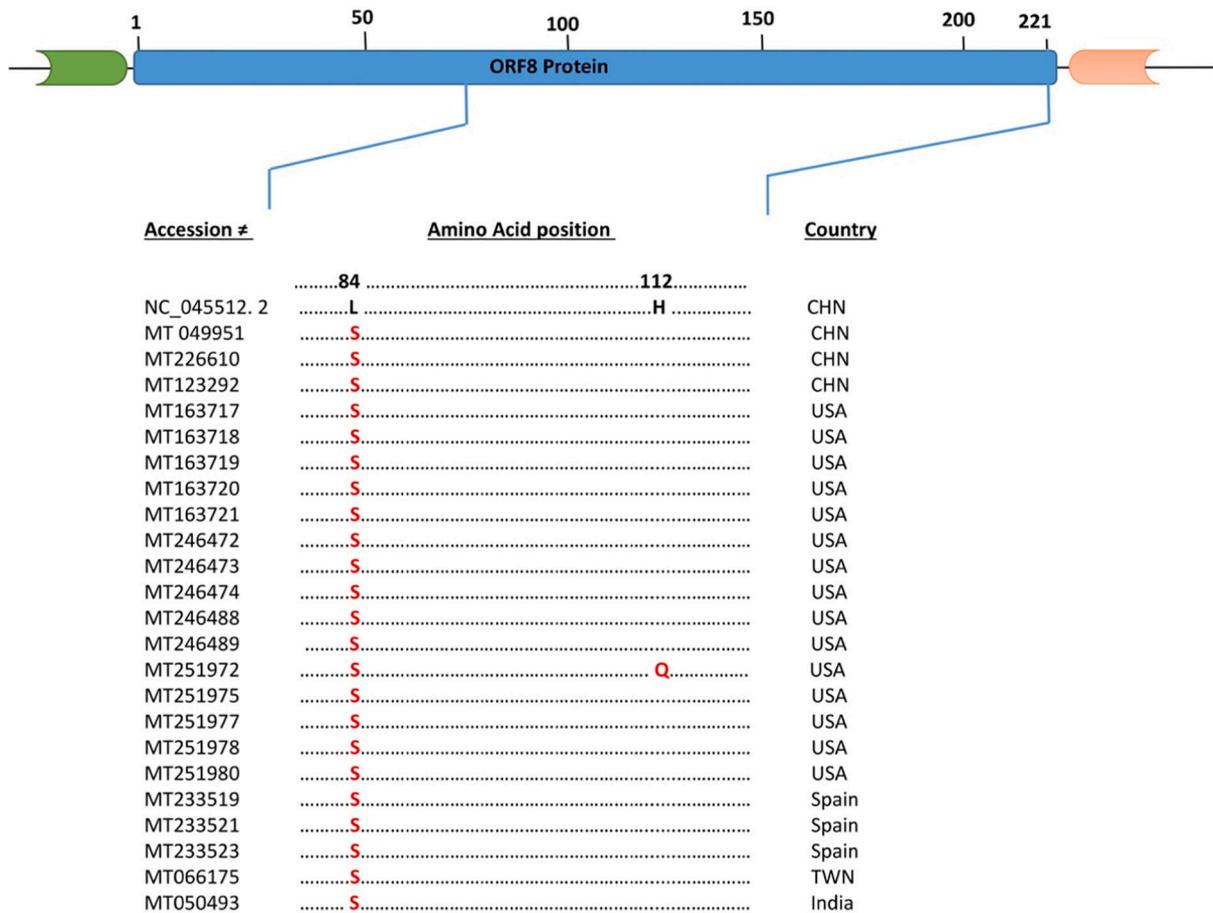


Fig. 4. Schematic diagram and Multiple Sequence Alignment of ORF8 protein of SARS-CoV-2. Dot represent identical amino acid and the letters highlighted in red represents substitution as compared to the reference sequence (NC\_045512.2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Furthermore, recent studies mentioned certain mutations at amino acid level in structural and accessory proteins of SARS-CoV-2 (Wang et al., 2020). In the present study the amino acids sequences of the four structural (S, N, E, and M) and ORF8 protein of 100 isolates from fifteen countries were analyzed for identification of mutations compared to the reference strain. The present study identified various substitutions of amino acids in the spike surface glycoprotein of different isolates, and one of the substitution at (aa 614 D to G) is accumulated in five different isolates from USA, while one Indian isolate have both deletion and substitution of amino acids as compared with reference sequence (NC\_045512.2) (Fig. 1). Interestingly the previous studies described that a single-amino-acid substitution of glycine (G) for glutamic acid (E) completely abolished the respiratory droplet transmission of Eurasian avian-like H1N1 (EAH1N1) swine influenza viruses (Zeng et al., 2017) and substitution of glycine (G) to aspartic acid (D) significantly attenuated H5N1 virus virulence in mice (Feng et al., 2015). In conclusion a glycine substitution may have important role in transmission, thus further studies are required to explore how these mutation may influence the transmission and tissue/host tropism of the SARS-CoV-2.

The analysis of N protein from different isolates revealed the presence of three different substitutions of amino acid in three locations. Furthermore, three Spanish isolates have substitution mutation at position aa197 S to L, one Chinese isolate at aa289 H to Q and one each Japanese and Chinese isolate have substitution at aa344 P to S (Fig. 2). Likewise, the present study identified mutation in M protein of only one isolate from USA (MT (163721) (Fig. 3). However, no mutation was found in the envelope protein which is supporting by the previous studies that indicated that E protein of SARS-CoV-2 is relatively conserved (Su et al., 2016). Interestingly, in ORF8 we found an accumulated substitution of Leucine, a nonpolar to Serine a polar amino acid at same position (aa84 L to S) in 23 different isolates as compared to the reference counterpart (NC\_045512.2) (Fig. 4). All the observed mutations in these isolates especially L to S could theoretically create a novel phosphorylation target for the mammalian host Serine/Threonine kinases of the host organism. And thus, can affect the conformation of the peptide. The results of the present study suggested that this virus is prone to mutations like most of other RNA viruses and with the passage of time it may become more potential threat to the world. Thus, further studies are required to characterize how these amino acids substitutions in different proteins affect the functionality and pathogenesis of SARS-CoV-2. Moreover, a continuous surveillance system is required to characterize the mutations undergone in this virus which are important for more effective vaccines development.

#### Declaration of competing interest

The authors declare that there is no conflict of interest.

#### References

Ceraolo, C., Giorgi, F.M., 2020. Genomic variance of the 2019-nCoV coronavirus. *J. Med.*

- Virol.* 92, 522–528.
- Chan, J.F., Lau, S.K., To, K.K., et al., 2015. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin. Microbiol. Rev.* 28, 465–522.
- Chan, J.F., 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, 221–236.
- Chen, N., Zhou, M., Dong, X., 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.
- China CDC, 2020. Tracking the epidemic. <http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm?from=timeline#Beijing%20Municipality%20Update>.
- Cui, J., Li, F., Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192.
- Drosten, C., Kellam, P., Memish, Z.A., 2014. Evidence for camel-to-human transmission of MERS coronavirus. *N. Engl. J. Med.* 371, 1359–1360.
- Feng, X., Wang, Z., Shi, J., et al., 2015. Glycine at position 622 in PB1 contributes to the virulence of H5N1 avian influenza virus in mice. *J. Virol.* 90, 1872–1879.
- Guan, Y.J., Zheng, B.J., He, Y.Q., et al., 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302, 276–278.
- Guan, W.J., Ni, Z., Hu, Y., et al., 2020. Clinical characteristics of 2019 novel coronavirus infection in China. *N. Engl. J. Med.* 382, 1708–1720.
- Hu, D., Zhu, C., Ai, L., et al., 2018. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg. Microbes Infect.* 7, 154–158.
- Huang, C., Wang, Y., Li, X., et al., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506.
- Kandeel, M., Al-Taher, A., Li, H., et al., 2018. Molecular dynamics of Middle East Respiratory Syndrome Coronavirus (MERS CoV) fusion heptad repeat trimers. *Comp. Biol. Chem.* 75, 205–212.
- Li, F., 2016. Structure, function, and evolution of coronavirus Spike proteins. *Annu. Rev. Virol.* 3, 237–261.
- Lu, R., Zhao, X., Li, J., et al., 2020a. Genomic characterization and epidemiology of 2019 novel coronavirus: implications of virus origins and receptor binding. *Lancet* 395, 565–574.
- Lu, R., Zhao, X., Li, J., et al., 2020b. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574.
- Lucas, M., Karrer, U., Lucas, A., et al., 2001. Viral escape mechanisms—escapology taught by viruses. *Int. J. Exp. Pathol.* 82, 269–286.
- Song, Z., Xu, Y., Bao, et al., 2019. From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses* 11, E59.
- Su, S., Wong, G., Shi, W., et al., 2016. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 24, 490–502.
- Tang, Q., Song, Y., Shi, M., et al., 2015. Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. *Sci. Rep.* 5, 17155.
- Wang, C., Liu, Z., Chen, Z., et al., 2020. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J. Med. Virol.* 20, 1–8.
- Weiss, S.R., Leibowitz, J.L., 2011. Coronavirus pathogenesis. *Adv. Virus Res.* 81, 85–164.
- WHO, 2020. Coronavirus disease (COVID-19) dashboard. <https://covid19.who.int/>.
- Wu, F., Zhao, S., Yu, B., et al., 2020a. A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269.
- Wu, A., Peng, Y., Huang, B., et al., 2020b. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* 27, 325–328.
- Zeng, W., Huanliang, Y., Yan, C., et al., 2017. A single-amino-acid substitution at position 225 in hemagglutinin alters the transmissibility of Eurasian avian-like H1N1 swine influenza virus in guinea pigs. *J. Virol.* 91, 800–817.
- Zhu, Z., Zhang, Z., Chen, W., et al., 2018. Predicting the receptor-binding domain usage of the coronavirus based on kmer frequency on spike protein. *Infect. Genet. Evol.* 61, 183–184.
- Zhu, N., Zhang, D., Wang, W., et al., 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733.