

# Unique SARS-CoV-2 Variant Exhibiting Plenteous Missense Mutations in Structural and Nonstructural Genes

Tarig M.S. Alnour<sup>a, b, c, \*</sup>, Mohammad Fahad Ullah<sup>b, c, \*\*</sup>,  
Elmutuz H. Elssaig<sup>a, b, \*\*\*</sup>, and Eltayib H. Ahmed-Abakur<sup>a, b, c, \*\*\*\*</sup>

<sup>a</sup> Faculty of Medical Laboratory Science, Department of Microbiology and Immunology Alzaiem Alazhari University, Khartoum North, 11111 Sudan

<sup>b</sup> Department of Medical Laboratory Technology (FAMS), University of Tabuk, P.O. Box 741, Tabuk, 71411 Saudi Arabia

<sup>c</sup> Prince Fahad Research Chair, University of Tabuk, P.O. Box 741, Tabuk, 71411 Saudi Arabia

\*e-mail: telnour@ut.edu.sa

\*\*e-mail: m.ullah@ut.edu.sa

\*\*\*e-mail: eelssaig@ut.edu.sa

\*\*\*\*e-mail: eosman@ut.edu.sa

Received April 12, 2021; revised April 28, 2021; accepted September 8, 2021

**Abstract**—Determining the variations in SARS-CoV-2 variant is considered main factor for understanding the pathogenic mechanisms, aid in diagnosis, prevention and treatment. The present study aimed to determine the genetic variations of SARS-CoV-2. The sequences of SARS-CoV-2 were obtained from National Center for Biotechnology Information (NCBI) and studied according to the time of isolation and their origin. The genome sequence of SARS-CoV-2 accession number NC\_045512 which represented the first isolated sequence of SARS-CoV-2 (Wuhan strain) was used as the reference sequence. The obtained genome sequences of SARS-CoV-2 were aligned against this Wuhan strain and variations among nucleotides and proteins were examined. The sequence of SARS-CoV-2 accession number MT577016 showed very low homology 98.75% compared to Wuhan strain NC\_045512. The analysis identified 301 nucleotide changes, which correspond to 258 different mutations; most of them 80% (207/258) were missense point mutations followed by 17.1% (44/258) silent point mutations. The critical mutations occurred in viral structural genes; 16.7% (43/258) mutations reported in S gene and 1 missense mutation was observed in E gene. Our finding showed the lowest homology and relatively distant phylogenetic relation of this SARS-CoV-2 variant with Wuhan strain along with high frequency of mutations including those in spike S and envelope E genes.

**Keywords:** COVID-19, genetic variation, homology, phylogenetic, SARS-CoV-2

**DOI:** 10.3103/S0095452721060153

## 1. INTRODUCTION

Six human coronaviruses had been identified before the emergence of SARS-CoV-2 (Raza et al., 2020; Wang et al., 2020) SARS-CoV-2 caused COVID-19 infection which is a more pathogenic form in comparison to the earlier identified SARS-CoV (2002) and Middle East respiratory syndrome coronavirus (MERS-CoV, 2013) (Naqvi et al., 2020). CoVs is enveloped, single strand and positive-sense RNA virus, have the largest RNA viral genome, (Lokman et al., 2020; Lu et al., 2020; Naqvi et al., 2020). The genome size of the SARS-CoV-2 varies from 29.8kb to 29.9 kb and followed the same pattern of gene characteristics of known CoVs (Khailany et al., 2020). The first study of full length genomic sequence of SARS-CoV-2 was done in China by Yongzhen Zhang team (Wu et al., 2020); it revealed that SARS-CoV-2 encodes 27 proteins from 14 ORFs including 15 non-structural, 4 major structural and

8 accessory protein. Spike glycoprotein (S), membrane (M), envelope (E) and nucleocapsid (N) are the four major structural proteins of SARS-CoV-2 (Lokman et al., 2020; Naqvi et al., 2020; Wang et al., 2020), the products of these structural genes play important roles in viral pathogenicity (Lokman et al., 2020). The accessory proteins are encoded by ORF8, ORF7a, ORF7b, ORF6 and ORF3a genes (Khailany et al., 2020).

However, RNA viruses tend to harbor error prone RNA dependent RNA polymerases which make the occurrence of mutations and recombination events rather frequent, this might play a role in the evolution of SARS-CoV-2. A recent study using phylogenetic network analysis has shown that the virus appears to be evolving into three distinct clusters; A and C being found mostly in Europe and America along while B being most common type in East Asia (Uddin et al., 2020). The study on genomic variation of SARS-CoV-2

is very important to analyze the disease course, pathogenesis, diagnosis, treatment and prevention (Wang et al., 2020), moreover it gives insights into the pattern of spread, genetic diversity and the dynamics of evolution (Khailany et al., 2020). The present study investigated the molecular variations of Telangana CoVID-19 variant MT577016.

## 2. METHODS

This study is an *in Silico* case report study; it is part of a project aimed to study the genetic variations of SARS-CoV-2 isolated in India; the study covered published sequences over a period of six months from March 2020 to September 2020. The genome sequences of SARS-CoV-2 were obtained from National Center for Biotechnology Information (NCBI) Virus Variation Resource repository (<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>) and these strains had been determined according to the time of isolation and their origin. The sequence of SARS-CoV-2 accession number NC\_045512 which represented the original Wuhan strain was used as the standard sequence. Genome sequences of SARS-CoV-2 was aligned against the Wuhan strain. The study was done using NCBI Nucleo-BLAST (NCBI) and variations of the nucleotides and proteins were stated. During collection of the data for this project; the sequence of SARS-CoV-2 accession number MT577016 showed very low homology when compared with Wuhan reference strain, other variants isolated from the same geographical area (India) and global variants therefore the changes in the nucleotides and mutations of this genome was determined.

Evolutionary relationships: the evolutionary relationship was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches (Felsenstein, 1985). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. This analysis involved 30 nucleotide sequences of SARS-2. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 29903 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018).

## 3. RESULTS

The SARS-CoV-2 accession number MT577016 which is the subject of this study was submitted by Yadav et al., 2020 on 14 March 2020 to Maximum Containment Laboratory, National Institute of Virology, Pashan, Pune, Maharashtra (India) and it got the accession number on 21 September 2020 (<https://www.ncbi.nlm.nih.gov/nuccore/MT577016.1?report=GenBank>). This iso-

late was named Telangana CoVID-19 variant in this study, as it was isolated in Telangana (India).

All the sequences of SARS-CoV-2 isolated in India showed high similarity (99.95–99.99%) to Wuhan strain NC\_045512 except Telangana CoVID-19 variant which showed 98.75% homology. The Telangana CoVID-19 variant showed 98.75% homology compared to the Wuhan strain NC\_045512. The analysis identified 301 nucleotide changes; most frequent nucleotide changes were T: A 120(39.9%), followed by C:A 62(20.6%) and T: G 53(17.6%) whereas the less frequent nucleotide changes showed in G:T and C:T (Fig. 1). These nucleotide change correspond to 258 different mutations; most of them 80.6% (208/258) were missense point mutations (Figs. 2a, 2b), followed by 17.1% (44/258) silent point mutations. The majority of these mutations 76.4% (197/258) occurred in the open reading frame 1 a/b (ORF1 a/b) while the ORF3a showed 4.7% (12/258) and few mutations 1.6% (4/258) occurred in 3'UTR terminal loop. The critical mutations occurred in viral structural genes; 16.7% (43/258) mutations reported in S gene; 8 were silent and 34 missense. One missense mutation was observed in E gene, whereas no mutation was detected in membrane (M) gene and nucleocapsid (N). Table 1 showed the mutations associated with the Telangana variant MT577016. As presented in Figs. 2a, 2b, the majority of the missense mutations that resulted in the change in codons are associated with the ORF1ab and relate to the changes in the corresponding amino acids encoded for several proteins such as leader protein, 3C-like proteinase, RNA-dependent RNA polymerase, helicase, 3'-5' exonuclease, endoRNase, 2'-O-ribose methyl transferase.

The Phylogenetic tree showed that Telangana CoVID-19 variant was located in separate node whereas all other SARS-CoV-2 variants clustered together closely in comparative analysis with Indian (Fig. 3a) and global variants (Fig. 3b), the accession number of each sequence was mentioned. Number at nodes indicates percent bootstrap value above 50 supported by more than 1000 replicates. The bar indicates the Jukes-Cantor evolutionary distance.

## 4. DISCUSSION

SARS-CoV-2, like other coronaviruses, contains a nonstructural gene with proofreading activity (Deng et al., 2020). As a typical RNA virus the average evolutionary rate for Co-Vs could be 10<sup>-4</sup> substitute per bp per year during each replication cycle (Ahmed-Abakura, 2020; Wang et al., 2020) However, knowledge of genomic interindividual variability and genetic variations could explain the discrepancies of spread, severity, and mortality of COVID-19 (Ahmed-Abakur and Alnour, 2020; Wang et al., 2020). As reported the present study determined the comparative genomic variations associated with the Telangana CoVID-19 variant.

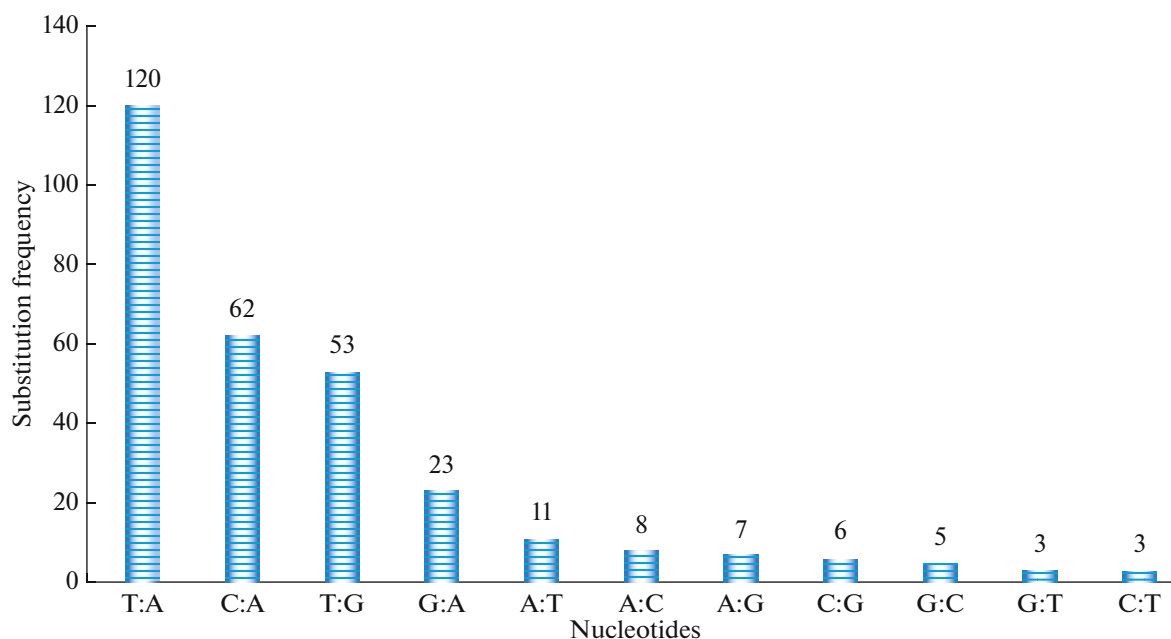


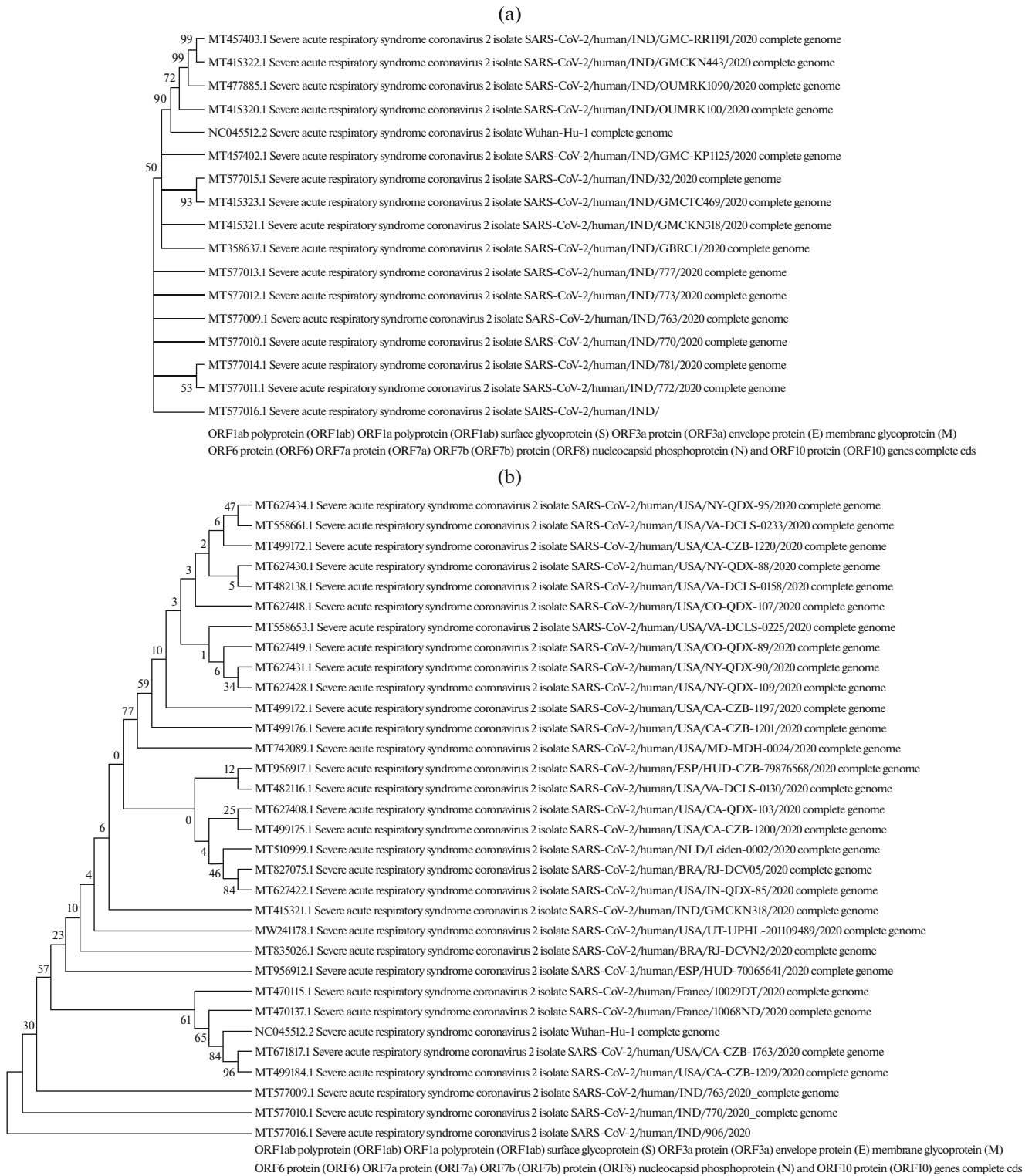
Fig. 1. Showed the frequency and types of nucleotide changes.

The genomic sequences of SARS-CoV-2 had been released by the worldwide scientific community in the last months to understand the molecular characteristics and evolutionary origin of this virus. Several authors determined the genetic variations of SARS-Co-19, and contrary to our findings all of them reported high homology. The Telangana CoVID-19 variant showed the lowest homology 98.75% compared to the published data and revealed high incidence mutations (302 nucleotides were changed and 258 different mutations were detected). Khailany et al., 2020 analyzed 95 SARS-CoV-2 complete genomes and reported only 116 mutations. While among 66 SARS-CoV-2 variants Ahmed-Abakur and Alnour 2020 reported 143 mutations and showed one variant as hyper mutated variant as it displayed 28 different point mutations. Further, Lu et al., 2020 studied 10 genome sequences of 2019-nCoV and showed more than 99.98% homology whereas Wang et al., 2020 analyzed 95 full-length genomic sequences of SARS-

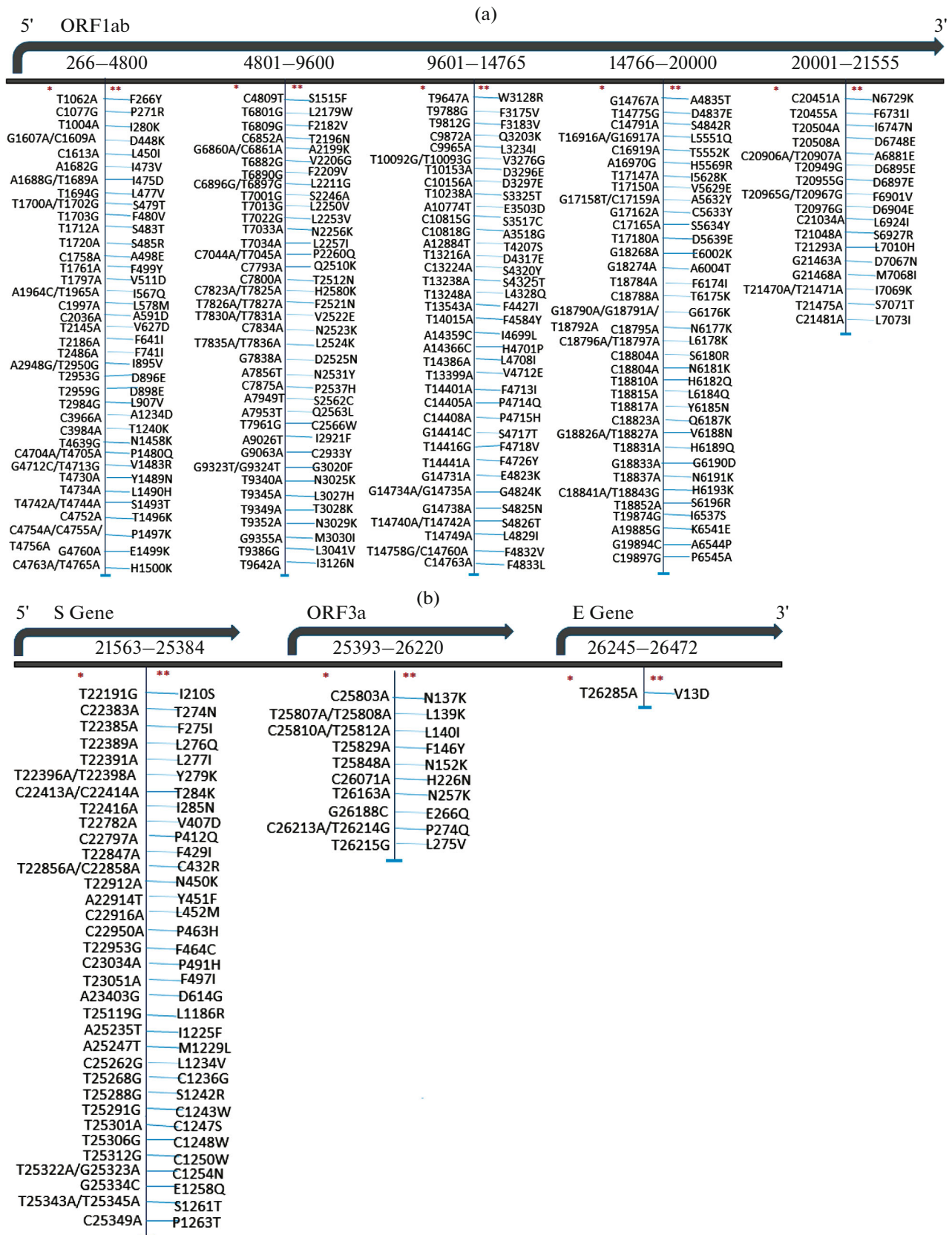
CoV-2 and reported that the homology among different isolates was extremely high (99.91%-100%) at the nucleotide level. In addition, Ceraolo and Giorgi, 2020 studied 56 SARS-CoV-2 genomes and showed high level of conservation (>99% sequence identity) and pointed only two core positions of high variability, one a silent variant in the ORF1ab and the other in ORF8 which resulted in two variants, ORF8-L and ORF8-S. Recently new variant of SARS-CoV-2 (VUI 202012/01) characterized with multiple spike protein mutations has been identified in United Kingdom that lead to huge increase in COVID-19 cases (with an estimated rise in reproductive number (R) by 0.4) (European Centre for Disease Prevention and Control, 2020). Thus the Telangana Covid 19 variant could be considered unique variant of SARS-2 based on sequence identity, frequency and occurrence of mutations, particularly as it displayed high incidence of mutations on structural genes which may affect the structural of the structural proteins and subsequently

Table 1. Distribution and site of mutations among Telangana variant MT577016

Site & type of mutation	5'UTR	ORF1ab	S gene	ORF3a	E gene	3' UTR	Total
5'UTR	1	0	0	0	0	0	1
Silent	0	34	8	2	0	0	44
Missense	0	162	34	10	1	0	207
Nonsense	0	1	0	0	0	0	1
Unknown	0	0	1	0	0	0	1
3' UTR	0	0	0	0	0	4	4
Total	1	197	43	12	1	4	258



**Fig. 2.** (a) Showed the landscape (ORF1ab) of Telangana CoVID-19 (MT577016) genome representing amino acid changes. The missense mutations (\*) resulting into corresponding amino acid changes (\*\*) in respective protein as marked. (b) Showed the landscape (S gene, ORF3a and E gene) of Telangana CoVID-19 (MT577016) genome representing amino acid changes. The missense mutations (\*) resulting into corresponding amino acid changes (\*\*) in respective protein as marked.



**Fig. 3.** (a) Showed the phylogenetic affiliation of Telangana CoVID-19 (MT577016) variant and sequences of closest phylogenetic neighbors which were retrieved from India. (b) Showed the phylogenetic affiliation of Telangana CoVID-19 (MT577016) variant and sequences of closest phylogenetic neighbors which were retrieved from different countries.

the pathogenicity. However, the emergence of the new variant could be attributed to; prolonged infection which may lead to accumulation of immune escape mutations at high rate, virus adaptation processes that occurs in susceptible animal species and then could transmitted back to humans (European Centre for Disease Prevention and Control, 2020).

Our study indicated that most of mutation occurred on ORF1 followed by S gene; these findings were in alignment with many reports in term of site of mutations (Ahmed-Abakur and Alnour, 2020; Khailany et al., 2020). However spike glycoprotein play important role during the entry of coronaviruses into host cells (Chang et al., 2012; Lokman et al., 2020; van Pesch et al., 2020; Shu and Gong, 2016), therefore any mutation in this gene might change the pattern of pathogenicity. It is understood that the hotspot mutations have the ability to cause changes in the amino acid sequences (Wang et al., 2020), resulting in significant changes in the stability, favoring various interactions, and conformational diversity (40). Our finding agreed with Khailany et al. 2020 where they did not detected mutations in N and M genes and disagreed with Wang et al., 2020 who stated that SARS-COV-2 is relatively conserved, especially in the E gene.

## CONCLUSIONS

This study presented evidence of a reportedly existing SARS-CoV-2 variant, which has shown an extraordinary ability to mutate even at the structural genes. Further studies are required to elucidate the exact role of genetic variations in SARS-CoV-2 that has challenged the human race, unprecedented in the recent history.

## FUNDING

This research is NOT funded fully or partially from any Governmental or non-governmental agencies.

## COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflict of interests. This article does not contain any studies involving animals or human participants performed by any of the authors.

## REFERENCES

- Ahmed-Abakura, E.H., Challenge of COVID 19: pathogenicity, genetic variations and laboratory diagnosis, *AJBSR*, 2020, vol. 11, no. 1. <https://doi.org/10.34297/AJBSR.2020.11.001604>
- Ahmed-Abakur, H.E. and Alnour, T.M.S., Genetic variations among SARS-CoV-2 strains isolated in China, *Gen. Rep.*, 2020, vol. 21, p. 100925. <https://doi.org/10.1016/j.genrep.2020.100925>
- Ceraolo, C. and Giorgi, F.M., Genomic variance of the 2019-nCoV coronavirus, *J. Med. Virol.*, 2020, vol. 92, no. 5, pp. 522–528. <https://doi.org/10.1002/jmv.25700>
- Chang, H.W., Egberink, H.F., Halpin, R., et al., Spike protein fusion peptide and feline coronavirus virulence, *Emerg. Infect. Dis.*, 2012, vol. 18, no. 7, pp. 1089–1095. <https://doi.org/10.3201/eid1807.120143>
- Deng, X., Gu, W., Federman, S., et al., Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California, *Science*, 2020, vol. 369, no. 6503, pp. 582–587. <https://doi.org/10.1126/science.abb9263>
- European Centre for Disease Prevention and Control, *Rapid Increase of a SARS-CoV-2 Variant with Multiple Spike Protein Mutations Observed in the United Kingdom 20 December 2020*, Stockholm: ECDC, 2020. <https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf>.
- Felsenstein, J., Confidence limits on phylogenies: an approach using the bootstrap, *Evolution*, 1985, vol. 39, pp. 783–791. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x>
- Khailany, R.A., Safdar, M., and Ozaslan, M., Genomic characterization of a novel SARS-CoV-2, *Gene Rep.*, 2020, vol. 19, p. 100682. <https://doi.org/10.1016/j.genrep.2020.100682>
- Kumar, S., Stecher, G., Li, M., et al., MEGA X: Molecular evolutionary genetics analysis across computing platforms, *Mol. Biol. Evol.*, 2018, vol. 35, no. 6, pp. 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Lokman, S.M., Rasheduzzaman, Salauddin, A., et al., Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach, *Infect. Genet. Evol.*, 2020, vol. 84, p. 104389. <https://doi.org/10.1016/j.meegid.2020.104389>
- Lu, R., Zhao, X., Li, J., et al., Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet*, 2020, vol. 395, no. 10224, pp. 565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
- Naqvi, A.A.T., Kisa, F., Taj, M., et al., Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach, *Biochim. Biophys. Acta Mol. Basis Dis.*, 2020, vol. 1866, no. 10, art. 165878. <https://doi.org/10.1016/j.bbadis.2020.165878>
- Raza, H., Wahid, B., Rubi, G., et al., Molecular epidemiology of SARS-CoV-2 in Faisalabad, Pakistan: a real-world clinical experience, *Infect. Genet. Evol.*, 2020, vol. 84, art. 104374. <https://doi.org/10.1016/j.meegid.2020.104374>
- Saitou, N. and Nei, M., The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.*, 1987, vol. 4, pp. 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Shu, B. and Gong, P., Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, vol. 113, no. 28, art. E4005–14. <https://doi.org/10.1073/pnas.1602591113>

- Tamura, K., Nei, M., and Kumar, S., Prospects for inferring very large phylogenies by using the neighbor-joining method, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, vol. 101, no. 30, pp. 11030–11035.  
<https://doi.org/10.1073/pnas.0404206101>
- Uddin, M., Mustafa, F., Rizvi, T.A., et al., SARS-CoV-2/COVID-19: viral genomics, epidemiology, vaccines, and therapeutic interventions, *Viruses*, 2020, vol. 12, no. 5, p. 526.  
<https://doi.org/10.3390/v12050526>
- van Pesch, V., van Eyll, O., and Michiels, T., The leader protein of Theiler's virus inhibits immediate-early alpha/beta interferon production, *J. Virol.*, 2001, vol. 75, no. 17, pp. 7811–7817.  
<https://doi.org/10.1128/jvi.75.17.7811-7817.2001>
- Wang, C., Liu, Z., Chen, Z., et al., The establishment of reference sequence for SARS-CoV-2 and variation analysis, *J. Med. Virol.*, 2020, vol. 92, no. 6, pp. 667–674.  
<https://doi.org/10.1002/jmv.25762>
- Wu, F., Zhao, S., Yu, B., et al., A new coronavirus associated with human respiratory disease in China, *Nature*, 2020, vol. 579, no. 7798, pp. 265–269.  
<https://doi.org/10.1038/s41586-020-2008-3>