



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.



Rare mutations in the accessory proteins ORF6, ORF7b, and ORF10 of the SARS-CoV-2 genomes

Sk. Sarif Hassan^{a,*}, Pabitra Pal Choudhury^b, Bidyut Roy^c

^a Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram 721140, India

^b Applied Statistics Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India

^c Human Genetics Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India

ARTICLE INFO

Keywords:

ORF6
ORF7b
ORF10
Rare mutations
Pathogenicity

ABSTRACT

A total number of 3080 SARS-CoV-2 genomes from all continents are considered from the NCBI database. Every accessory protein ORF6, ORF7b, and ORF10 of SARS-CoV-2 possess a single missense mutation in less than 1.5% of the 3080 genomes. It has now been observed that different non-synonymous mutations occurred in these three accessory proteins. Most of these rare mutations are changing the amino acids such as hydrophilic to hydrophobic, acidic or basic to hydrophobic, and vice versa etc. So these highly conserved proteins might play an essential role in virus pathogenicity. This study opens a question whether it carries some messages about the virus rapid replications, and virulence.

1. Introduction

The accessory proteins of SARS-CoV-2 help the virus to infect its human host, replicate, and eventually spread from person to person. There are six accessory proteins ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 in the SARS-CoV-2 genomes, which play important roles in the viral life cycle, and may contribute to the pathogenesis and virulence (Xu et al., 2009). These accessory proteins are unique to SARS-CoV-2, as they have least homology in amino acid sequence with accessory proteins of other coronaviruses. These proteins contribute to virus virulence though it does not directly affect the virus release, stability and pathogenesis (Xu et al., 2009). Therefore, close observations and understanding of these proteins may explain the differences in pathogenicity between SARS-CoV-2, and other known coronaviruses. Within a very quick time span, SARS-CoV-2 virus has been evolving for accounting various missense mutations in various proteins. It has been already observed by us that except one in ORF7b, there are no missense mutations over the proteins ORF6, and ORF10 across the SARS-CoV-2 genomes detected in 128 Indian patients (*Missense mutations in sars-cov2 genomes from indian patients*, 2020; Hassan et al., 2020). Missense mutations in the accessory proteins of SARS-CoV-2 might impact in heterogeneity of SARS-CoV-2 and may contribute to comprehending the pathogenic dynamics as well as virulence of the virus over time. The amino acid mutations among SARS-CoV-2 isolates from diverse

locations could be linked with their geographical distributions. This rapidly evolving virus is capable of adapting swiftly to the diverge environments (Islam et al., 2020). So it is quite important to discover whether any missense mutations are found in those proteins across the SARS-CoV-2 genomes worldwide, other than India.

1.1. Functions of the accessory proteins ORF6, ORF7b, and ORF10

Here we present the some known reported functions of three accessory proteins ORF6, ORF7b, and ORF10 as follows:

- **ORF6:** This protein contains a hydrophobic N-terminal, and it has been suggested to have a N-endo-C-endo conformation (Gunalan et al., 2011). The ORF6 protein makes intracellular membrane rearrangements resulting in a vesicular population in the infected cell which could possibly serve some role in increasing replication of the virus (Frieman et al., 2007). It has been previously reported that ORF6 down-regulated the mRNA level of the co-transfected myc-nsp8 gene (Zhou et al., 2010). It is also reported that amino acids 53–56 in ORF6 of SARS-CoV constitute a putative diacidic motif (DDEE) which influences the suppression of the expression of co-transfected myc-nsp8 (Gunalan et al., 2011).
- **ORF7b:** ORF7b is very much unique since it does not have any sequence homology with other viral proteins (Schaecher et al., 2007). The accessory protein ORF7b has been reported to have a structural

* Corresponding author.

E-mail addresses: sarimif@gmail.com (Sk.S. Hassan), broy@isical.ac.in (B. Roy).

Table 1

Missense mutations in ORF6, ORF7b, and ORF10 across 3080 SARS-CoV-2 genome sequences.

Proteins	Geo-location	Mutations	R-group property	Proteins	Geo-location	Mutations	R-group property
QJC19423-ORF6	Germany	V(24)A	Hydrophobic to Hydrophobic	QJY40403-ORF7b*	India-Botad	S(31)L	Hydrophilic to Hydrophobic
QJT72174-ORF6	France	D(61)Y	Acidic to Hydrophilic	QJX70209-ORF7b	USA-MI	T(40)I	Hydrophilic to Hydrophobic
QJT72858-ORF6	France	H(3)Y	Basic to Hydrophilic	QJX74553-ORF7b	USA	L(20)STOP	Hydrophobic to STOP
QJS54110-ORF6	Greece: Athens	T(21)I	Hydrophilic to Hydrophobic	QJS54820-ORF7b	USA:CA	F(30)L	Hydrophobic to Hydrophobic
QJS56575-ORF6	USA:WA	I(33)T	Hydrophobic to Hydrophilic	QJQ84777-ORF7b	THILAND	C(41)F	Hydrophilic to Hydrophobic
QJR84997-ORF6	USA:CA	D(53)G	Acidic to Hydrophobic	QJQ84801-ORF7b	THILAND	C(41)F	Hydrophilic to Hydrophobic
QJR87301-ORF6	Australia: Victoria	D(6)Y	Acidic to Hydrophilic	QJQ84813-ORF7b	THILAND	C(41)F	Hydrophilic to Hydrophobic
QJR87841-ORF6	Australia: Victoria	W(27)L	Hydrophobic to Hydrophobic	QJQ84825-ORF7b	THILAND	C(41)F	Hydrophilic to Hydrophobic
QJR89461-ORF6	Australia: Victoria	D(53)Y	Acidic to Hydrophilic	QJQ84837-ORF7b	THILAND	C(41)F	Hydrophilic to Hydrophobic
QJR91693-ORF6	Australia: Victoria	W(27)L	Hydrophobic to Hydrophobic	QJD47604-ORF7b	USA:CT	F(19)L	Hydrophobic to Hydrophobic
QJQ39516-ORF6	USA:MI	I(33)T	Hydrophobic to Hydrophilic	QJC19833-ORF7b	USA:WA	F(28)Y	Hydrophobic to Hydrophilic
QJD48547-ORF6	USA:WA	K(42)N	Basic to Hydrophilic	QKE45662-ORF7b	USA:CA	T(40)I	Hydrophilic to Hydrophobic
QJD49135-ORF6	USA:WA	K(42)N	Basic to Hydrophilic	QKE45866-ORF7b	USA:CA	F(30)L	Hydrophobic to Hydrophobic
QJD49147-ORF6	USA:WA	K(42)N	Basic to Hydrophilic	QJY51909-ORF10	USA:MI	R(24)C	Basic to Hydrophilic
QJC21021-ORF6	Spain	N(34)S	Hydrophilic to Hydrophilic	QJX70372-ORF10	USA:IL	I(4)L	Hydrophobic to Hydrophobic
QJA16692-ORF6	USA:WA	K(42)N	Basic to Hydrophilic	QJS56880-ORF10	USA:ID	S(23)F	Hydrophilic to Hydrophobic
QJA16812-ORF6	USA:WA	D(61)Y	Acidic to Hydrophilic	QJR96431-ORF10	USA:CA	Q(29)STOP	Acidic to STOP
QIU81889-ORF6	China: Beijing	Q(8)H	Hydrophilic to Basic	QJQ38873-ORF10	USA:CA	S(23)F	Hydrophilic to Hydrophobic
QIS60718-ORF6	USA:WA	K(42)N	Basic to Hydrophilic	QJD48984-ORF10	USA:WA	R(24)L	Basic to Hydrophobic
QIQ50036-ORF6	USA:WA	V(9)F	Hydrophobic to Hydrophobic	QIS60555-ORF10	USA:WA	I(4)L	Hydrophobic to Hydrophobic
QIQ50136-ORF6	USA:WA	K(42)N	Basic to Hydrophilic	QIS29991-ORF10	China: Hubei, Wuhan	V(6)I	Hydrophobic to Hydrophobic

* QJY40403-ORF7b from India-Botad. Note that mutations marked as 'bold' are reported here for the first time.

component in the SARS-CoV virions (Pekosz et al., 2006). The transmembrane domain of ORF7b is essential to retain the protein in the Golgi compartment (Schaecher et al., 2008). It has been found that ORF7b of SARS-CoV-2 is 81% identical to that of SARS-CoV. It is reported that absence of ORF7b does not affect virus replication (Pekosz et al., 2006). Alanine scanning experiments showed that the amino acids in 13–15, and 19–22 are critical for the retention of ORF7b in the Golgi complex (Schaecher et al., 2008).

• **ORF10:** ORF10 protein is the unique accessory protein in SARS-CoV-2 as it does not present in the SARS-CoV (Taiaroa et al., 2020), and hence it is sometimes called "mysterious protein". The signature appearance of ORF10 in SARS-CoV-2 makes the novel coronavirus easily detectable compared to other PCR based methods (Koyama et al., 2020). Function of this protein is still undiscovered. Therefore any missense mutation in ORF10 is significantly worth determining, and requires further investigations in order to understand the functions of ORF10.

2. Methods

All the protein sequences of 3080 SARS-CoV-2 genomes are fetched from the NCBI virus database. Then for each of the three accessory proteins, amino acid sequences are exported in fasta format using file operations through Matlab. For each accessory protein, these sequences (fasta formatted) are blasted using Clustal-Omega, and found the mutations and accordingly their associated positions are accounted (Madeira et al., 2019).

3. Results

It is to be noted that among these virus genomes from 3080 patients, there are 2126 genomes from USA, 306 from Asia, 281 genomes from Europe, 365 genomes from Oceania and one genome from Africa. Here (Table 1), we present the missense mutations of the three accessory proteins ORF6, ORF7b, and ORF10 over the available 3080 SARS-CoV-2 genomes. Note that the mutation of an amino acid A_1 to an amino acid A_2 is denoted by $A_1(l)A_2$ where l denotes location in the reference amino

Table 2

25 SARS-CoV-2 genomes from USA:CA having D(53)G mutation in ORF6.

USA: CA		
Mutation: D(53)G		Acidic to Hydrophobic
QKE49188-ORF6	QKE49296-ORF6	QKE49404-ORF6
QKE49200-ORF6	QKE49320-ORF6	QKE49416-ORF6
QKE49212-ORF6	QKE49332-ORF6	QKE49428-ORF6
QKE49224-ORF6	QKE49344-ORF6	QKE49440-ORF6
QKE49236-ORF6	QKE49356-ORF6	QKE49452-ORF6
QKE49248-ORF6	QKE49368-ORF6	QKE49464-ORF6
QKE49260-ORF6	QKE49380-ORF6	QKE49476-ORF6
QKE49272-ORF6	QKE49392-ORF6	QKE49488-ORF6
		QKE49500-ORF6

acid sequence. In Table 1, also the changes of the R-group of each amino acid according to the mutations are presented. The Table 1 depicts the following:

- In less than 1.5% of the SARS-CoV-2 genomes, the accessory proteins ORF6 (1.49%), ORF7b (0.422%), and ORF10 (0.259%) possess the missense mutations as mentioned in Table 1.
- The ORF7b protein (QJX74553) got truncated due to mutation L(20)STOP to a stop codon. Also the protein ORF10 (QJR96431) is truncated due to the mutation Q(29)STOP.
- There are several changes of the R-group property of amino acids viz. hydrophobic to hydrophobic, acidic to hydrophilic, basic to hydrophilic, hydrophilic to hydrophobic, acidic to hydrophobic, hydrophobic to hydrophilic, hydrophilic to basic and basic to hydrophobic.
- The putative diacidic motif (DDEE) in SARS-CoV of the ORF6 protein in the 53–56 position has been mutated to DEEQ in ORF6 in some of the SARS-CoV-2 genomes. This motif DEEQ of ORF6 has been mutated (D(53)Y) to YEEQ in QJR89461-ORF6 (Australia: Victoria). Strikingly, the motif DEEQ mutated (D(53)G) to GEEQ in the ORF6 protein across the 25 SARS-CoV-2 genomes from USA:CA as shown in Table 2. These missense mutations may affect the suppression of the expression of co-transfected myc-nsp8.
- The transmembrane domain of ORF7b in SARS-CoV-2 is fully

conserved with respect to that of SARS-CoV. In the virus sequence QJD47604-ORF7b (USA:CT) possesses a mutation F(19)L which might affect the retention of ORF7b in the Golgi complex. There is nonsense mutation L(20)STOP in the QJX74553-ORF7b (USA), which makes the protein non-functional.

- The nonsense mutation in QJR96431-ORF10 (USA:CA) makes the protein non-functional. In USA, three distinct missense mutations R(24)C, I(4)L and S(23)F are found in each pair of genomes viz. {QJY51909 – ORF10, QJD48984 – ORF10}, {QJX70372 – ORF10, QIS60555 – ORF10}, and {QJS56880 – ORF10, QJQ38873 – ORF10} respectively. Also there is another mutation V(6)I (hydrophobic to hydrophobic) in the virus QIS29991-ORF10 (China: Hubei, Wuhan) which might not affect the functions of ORF10 because of the synonymous change of R-group property.

4. Concluding remarks

The novel RNA virus SARS-CoV-2 has been evolving through several mutations over the associated proteins within a very short time from its discovery since December 2019. It has now been observed that few mutations occurred in these three proteins. As we see here, the total amount of mutations possessed by three accessory proteins ORF6, ORF7b, and ORF10 is around 1.5%. Since the mutations are few, we firmly believe that these mutated protein variants are beneficial for virus. This study opens a question whether it carries some messages about the virus rapid replications and virulence and their role in pathogenicity. Whether these mutations raise different symptoms in the affected patients, is certainly of worth investigation from the pathogenetic perspective. These conserved proteins may suggest for development of antiviral drugs or formulate new vaccines from a new starting point. Since the mutations are rare in numbers, we firmly believe that these mutated protein variants might elongate sustainability and pathogenicity of the virus. It will also be interesting to explore whether these rare mutations were evolved from China or other affected continents/countries. Recently, two variants VUI – 202012/01 and 501. V2 of SARS-CoV-2 (first detected in Europe and South Africa, respectively) became a concern to the world population due to the spread in different countries (Loconsole et al., 2021), GISAID. Of these two, VUI-202012/01 has more transmission ability. Sequence analysis has shown that these variants have mutations in S, N, NSP etc. genes and none of them has mutation in ORF6, ORF7b, and ORF10. So, it corroborates to our observation that these ORF proteins are highly conserved in the SARS-CoV2 genome. Although the clinical importance of these rare mutations are yet to be discovered, this study certainly would serve as a groundwork.

Author contributions

SH conceived the problem. SH determined the mutations. SH, PPC, and BR analysed the data and result. SH wrote the initial draft which was checked and edited by all other authors to generate the final version.

Declaration of Competing Interest

The authors do not have any conflicts of interest to declare.

References

- Frieman, M., Yount, B., Heise, M., Kopecky-Bromberg, S.A., Palese, P., Baric, R.S., 2007. Severe acute respiratory syndrome coronavirus orf6 antagonizes stat1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/golgi membrane. *J. Virol.* 81 (18), 9812–9824.
- Gunalan, V., Mirazimi, A., Tan, Y.-J., 2011. A putative diacidic motif in the sars-cov orf6 protein influences its subcellular localization and suppression of expression of co-transfected expression constructs. *BMC Res. notes* 4 (1), 446.
- Hassan, S.S., Choudhury, P.P., Basu, P., Jana, S.S., 2020. Molecular conservation and differential mutation on orf3a gene in indian sars-cov2 genomes, 112 (5), pp. 3226–3237.
- Islam, M.R., Hoque, M.N., Rahman, M.S., Alam, A.R.U., Akther, M., Puspo, J.A., Akter, S., Sultana, M., Crandall, K.A., Hossain, M.A., 2020. Genome-wide analysis of sars-cov-2 virus strains circulating worldwide implicates heterogeneity. *Sci. Rep.* 10 (1), 1–9.
- Koyama, T., Platt, D., Parida, L., 2020. Variant analysis of covid-19 genomes. *Bull. World Health Organ.* 98, 495–504.
- Loconsole, D., Sallustio, A., Accogli, M., Centrone, F., Capozzi, L., Del Sambro, L., Parisi, A., Chironna, M., 2021. Genome sequence of a sars-cov-2 vui 202012/01 strain identified from a patient returning from London, England, to the apulia region of Italy. *Microbiol. Res. Announc.* 10 (4).
- Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R., Potter, S.C., Finn, R.D., et al., 2019. The embl-ebi search and sequence analysis tools apis in 2019. *Nucleic Acids Res.* 47 (W1), W636–W641.
- Missense mutations in sars-cov2 genomes from indian patients, 2020.
- Pekosz, A., Schaecher, S.R., Diamond, M.S., Fremont, D.H., Sims, A.C., Baric, R.S., 2006. Structure, expression, and intracellular localization of the sars-cov accessory proteins 7a and 7b. In: *The Nidoviruses*. Springer, pp. 115–120.
- Schaecher, S.R., Mackenzie, J.M., Pekosz, A., 2007. The orf7b protein of severe acute respiratory syndrome coronavirus (sars-cov) is expressed in virus-infected cells and incorporated into sars-cov particles. *J. Virol.* 81 (2), 718–731.
- Schaecher, S.R., Diamond, M.S., Pekosz, A., 2008. The transmembrane domain of the severe acute respiratory syndrome coronavirus orf7b protein is necessary and sufficient for its retention in the golgi complex. *J. Virol.* 82 (19), 9477–9491.
- Taiaroa, G., Rawlinson, D., Featherstone, L., Pitt, M., Caly, L., Druce, J., Purcell, D., Harty, L., Tran, T., Roberts, J., et al., 2020. Direct rna sequencing and early evolution of sars-cov-2. *bioRxiv*.
- Xu, K., Zheng, B.-J., Zeng, R., Lu, W., Lin, Y.-P., Xue, L., Li, L., Yang, L.-L., Xu, C., Dai, J., et al., 2009. Severe acute respiratory syndrome coronavirus accessory protein 9b is a virion-associated protein. *Virology* 388 (2), 279–285.
- Zhou, H., Ferraro, D., Zhao, J., Hussain, S., Shao, J., Trujillo, J., Netland, J., Gallagher, T., Perlman, S., 2010. The n-terminal region of severe acute respiratory syndrome coronavirus protein 6 induces membrane rearrangement and enhances virus replication. *J. Virol.* 84 (7), 3542–3551.