



Article

Emergence of Unique SARS-CoV-2 ORF10 Variants and Their Impact on Protein Structure and Function

Sk. Sarif Hassan^{1,*} , Kenneth Lundstrom², Ángel Serrano-Aroca³, Parise Adadi⁴, Alaa A. A. Aljabali⁵, Elrashdy M. Redwan^{6,7}, Amos Lal⁸, Ramesh Kandimalla⁹, Tarek Mohamed Abd El-Aziz^{10,11}, Pabitra Pal Choudhury¹², Gajendra Kumar Azad¹³, Samendra Sherchan¹⁴, Murtaza Tambuwala¹⁵, Gaurav Chauhan¹⁶, Kazuo Takayama¹⁷, Debmalya Barh^{18,19}, Giorgio Palù²⁰, Pallab Basu²¹, Vladimir Uversky^{22,*} 

- ¹ Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram, Paschim Medinipur 721140, West Bengal, India; sksarifhassan@pinglacollege.ac.in
- ² PanTherapeutics, Rte de Lavaux 49, CH1095 Lutry, Switzerland; lundstromkenneth@gmail.com
- ³ Biomaterials and Bioengineering Lab, Centro de Investigación Traslacional San Alberto Magno, Universidad Católica de Valencia San Vicente Mártir, c/Guillem de Castro, 94, 46001 Valencia, Valencia, Spain; angel.serrano@ucv.es
- ⁴ Department of Food Science, University of Otago, Dunedin 9054, New Zealand; pariseadadi@gmail.com
- ⁵ Department of Pharmaceutics and Pharmaceutical Technology, Faculty of Pharmacy, Yarmouk University, Irbid 566, Jordan; alaa@yu.edu.jo
- ⁶ Biological Science Department, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia; lradowan@kau.edu.sa
- ⁷ Therapeutic and Protective Proteins Laboratory, Protein Research Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications, New Borg EL-Arab, 21934, Alexandria, Egypt
- ⁸ Department of Medicine, Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, Minnesota, USA; manavamos@gmail.com
- ⁹ Applied Biology, CSIR-Indian Institute of Chemical Technology, Uppal Road, Tarnaka, Hyderabad-500007, Telangana State, India; ramesh.kandimalla@gmail.com
- ¹⁰ Department of Cellular and Integrative Physiology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Dr, San Antonio, TX 78229-3900, USA; mohamedt1@uthscsa.edu
- ¹¹ Zoology Department, Faculty of Science, Minia University, El-Minia 61519, Egypt
- ¹² Applied Statistics Unit, Indian Statistical Institute, Kolkata 700108, India; pabitra@isical.ac.in
- ¹³ Department of Zoology, Patna University, Patna, Bihar, India; gkazad@patnauniversity.ac.in
- ¹⁴ Department of Environmental Health Sciences, Tulane University, New Orleans, LA 70112, USA; sshercha@tulane.edu
- ¹⁵ School of Pharmacy and Pharmaceutical Science, Ulster University, Coleraine BT52 1SA, Northern Ireland, UK; m.tambuwala@ulster.ac.uk
- ¹⁶ School of Engineering and Sciences, Tecnológico de Monterrey, Av. Eugenio Garza Sada 2501 Sur, 64849 Monterrey, Nuevo León, Mexico; gchauhan@tec.mx
- ¹⁷ Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto 606-8507, Japan; kazuo.takayama@cira.kyoto-u.ac.jp
- ¹⁸ Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, WB, India; dr.barh@gmail.com
- ¹⁹ Departamento de Genética, Ecología e Evolucao, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
- ²⁰ Department of Molecular Medicine, University of Padova, Via Gabelli 63, 35121, Padova, Italy; giorgio.palu@unipd.it
- ²¹ School of Physics, University of the Witwatersrand, Johannesburg, Braamfontein 721140, South Africa; pallabbasu@gmail.com
- ²² Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA; vuversky@usf.edu
- * Correspondence: vuversky@usf.edu; sksarifhassan@pinglacollege.ac.in.

Abstract: The devastating impact of the ongoing coronavirus disease 2019 (COVID-19) on public health, caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has made fighting of the COVID-19 pandemic a top priority in medical research and pharmaceutical development. Surveillance of SARS-CoV-2 mutations is essential for the comprehension of SARS-CoV-2 variant diversity and their impact on virulence and pathogenicity. The SARS-CoV-2 open reading frame 10 (ORF10) protein interacts with multiple human proteins CUL2, ELOB, ELOC,

MAP7D1, PPT1, RBX1, THTPA, TIMM8B, and ZYG11B expressed in the lung tissues. Mutations and co-mutations in the emerging SARS-CoV-2 ORF10 variants are expected to impact the severity of the virus and its associated consequences. In this article, We highlight 128 single mutations and 35 co-mutations in the unique SARS-CoV-2 ORF10 variants in this article. The possible predicted effects of these mutations and co-mutations on the secondary structure of ORF10 variants and host protein interactomes are presented. The findings highlight the possible effects of mutations and co-mutations on the emerging 140 ORF10 unique variants from secondary structure and intrinsic protein disorder perspectives.

Keywords: SARS-CoV-2, ORF10, Co-mutations, Intrinsic Protein Disorder, and Ubiquitin Ligase Complex

1. Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) continues the pandemic spread of coronavirus disease 2019 (COVID-19), with over 189 million people confirmed infected and at least 4.07 million deaths worldwide [1,2]. In 2021, in almost every region of the SARS-CoV-2 genome, several mutations compared to the wild SARS-CoV-2 (NC_045512) were discovered [3–5]. The SARS-CoV-2 is constantly evolving and new variants with different characteristics are emerging [6,7]. Detection and mutation surveillance of SARS-CoV-2 is of utmost priority investigate the origin and to combat the virus [8]. To date, no method can rapidly diagnose multiple viral infections and determine variants in a high-throughput manner [9]. SARS-CoV-2 is one of the largest RNA viruses with a genome of approximately 29Kb, which includes eleven open reading frames (ORFs) [10–13]. These ORFs (1a and 1b) possessed two polypeptides that are translated into sixteen non-structural proteins (NSP1-16) [14–16]. The main non-structural proteins (NSP) include RNA-dependent RNA polymerase (RdRp or NSP12) and a 3′-5′ exonuclease [17]. RNA viruses typically possess a high mutation rate ranging from 10^{-6} to 10^{-4} [18,19]. Notably, SARS-CoV-2 possesses the 3′-5′ exonuclease capable of correcting mistakes during replication [20]. While surveilling mutations to comprehend the genetic diversity across various SARS-CoV-2 variants, it is also important to decipher whether the increase in mutation frequency is because of the natural selection, and to determine the possible consequences for SARS-CoV-2 fitness, such as increased infectivity and pathogenicity, or due to adaptation, thereby becoming drug-resistant or possessing ability to evade the immune system [21,22]. Non-synonymous mutations of various SARS-CoV-2 proteins have been reported [23–26].

The ORF proteins are dispensable for viral growth *in vitro*, and might play an important roles within the environment of the infected host [27]. The SARS-CoV-2 open reading frame 10 (ORF10) protein shows no sequence similarity with other known coronavirus protein [28–30]. The SARS-CoV-2 ORF10, a putative 38-amino acid viral protein encoded in the 3′ accessory region of the genome, is a highly ordered, hydrophobic, and thermally stable protein, which contains at least one transmembrane region [29,31]. The ORF10 binds to components of a Cullin-2-RING-ligase (CRL2) complex containing Cullin-2, RBX1, Elongin B, Elongin C, and ZYG11B ($CRL2^{ZYG11B}$) [32–34]. Earlier, it has been reported that the extreme N terminus of ORF10 contains a methionine-glycine-tyrosine motif, which would presumably aid ORF10 to be recruited into $CRL2^{ZYG11B}$ ubiquitin ligase complex [33]. It was further confirmed that interaction between ORF10 and $CRL2^{ZYG11B}$ is not relevant for SARS-CoV-2 infection *in vitro* [30,33]. No evidence was observed whether ORF10 regulates or is regulated by $CRL2^{ZYG11B}$ [33]. On the other hand, ORF10 protein in some SARS-CoV-2 variants, resulted in non-attenuation of disease and maintained transmissibility [35]. Furthermore, ORF10 is not essential for viral infection and replication while encoding a truncated protein that is neutrally

evolving, through positive selection [30,36].

In this study, we report 128 single mutations and 35 co-mutations in the unique SARS-CoV-2 ORF10 variants. This report illuminates potential effects due to mutations and co-mutations in the emerging ORF10 variants from the secondary structure and intrinsic protein disorder perspectives.

2. Materials and Methods

2.1. Data

A total of 202968 SARS-CoV-2 ORF10 sequence data were retrieved from the National Center for Biotechnology Information (NCBI) database on June 29, 2021. Note that none of these complete ORF10 sequences contain any ambiguous characters. Among 202968 SARS-CoV-2 ORF10 sequences only 140 sequences were unique and distinct, globally. Furthermore, SARS-CoV-2 ORF10 protein sequences from the GISAID database were considered for finding co-mutations from the CoVal database.

2.2. Methods

2.2.1. Transmembrane topology and secondary structure prediction

Prediction of a transmembrane protein topology of a given protein is one of the classical issues in bioinformatics. The Phobius program was used to predict transmembrane topology for SARS-CoV-2 ORF10 protein variants [37,38].

The secondary structure of the SARS-Cov-2 ORF10 protein was predicted using the JPred 4 webserver [39]. In addition to protein secondary structure JPred was used to predict solvent accessibility and coiled-coil regions. The following keys were used in the secondary structure prediction of ORF10 variants [39].

- Shades of red: The more red a position is, the higher the level of conservation of chemical properties of the amino acids.
- Jnetpred: Final secondary structure prediction for a query.
- Jnet_25: Jnet prediction of burial, less than 25% solvent accessibility.
- Jnet_5: Jnet prediction of burial, less than 5% exposure.
- Jnet_0: Jnet prediction of burial, 0% exposure.
- Jnetconf: Jnet reliability of prediction accuracy ranges from 0 to 9, bigger is better.

2.2.2. Intrinsic disorder analysis

All SARS-CoV-2 ORF10 variants were subjected to the per-residue disorder analysis, with PONDR-VSL2 algorithm [40]. This tool showed good performance on proteins containing both structure and disorder, and is considered as an accurate standalone disorder predictor [41–43]. Predisposition scores for the per-residual conditions are 0 to 1, where 0 indicates residues entirely arranged, and 1 indicates residues completely disordered. Residues with disorder scores between 0.25 and 0.5 were considered as disordered, residues with disorder scores between 0.1 and 0.25 were considered moderately disordered, whereas residues with values higher than 0.5 were considered disordered.

2.2.3. Analysis of sequence variation

Single mutations in all the 140 unique ORF10 proteins were determined using the Virus Pathogen Resource ViPR by uploading a Fasta file of ORF10 sequences [44]. A snapshot of the ORF10 sequence variations is presented in Figure 1.

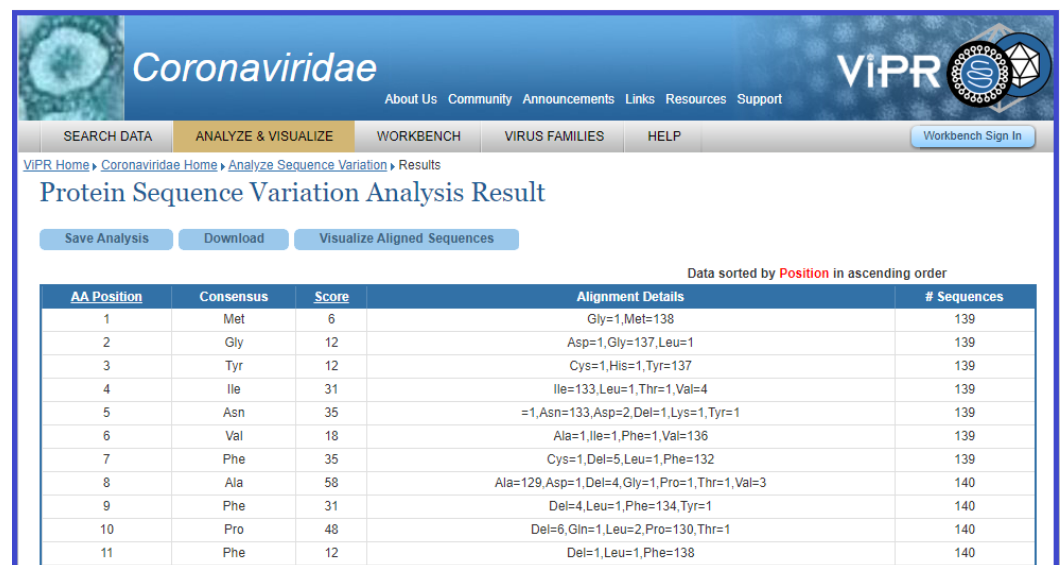


Figure 1. A snapshot of ViPR database showing the ORF10 sequence variations and consensus amino acid residues.

Furthermore, the predicted effect on pathogenicity of all the mutations was analyzed with [PredictSNP](#) and [PhD-SNP](#) [45,46]. In addition, co-occurrence of mutations in ORF10 proteins were also detected by the [CoVal](#) database.

A quantitative measure was defined to localize the co-occurrent mutations in an ORF10 protein in a given geo-location with regards to other geo-locations worldwide.

Localization across countries: For each co-occurrent mutation in an ORF10 protein, localization is defined as

$$NF_m = \frac{NG_m}{TG_m}$$

Where, NG_m and TG_m denote the number of SARS-CoV-2 genomes with these specific simultaneous mutations (m) in a geo-location, and the total number of SARS-CoV-2 genomes with this mutation (m) worldwide, respectively. It varies from 0 to 1. The normalized factor 0 denotes uniformly spreading of the mutations across various geo-locations, whereas 1 denotes the detection of the mutation in a single geo-location (discussed in [CoVal](#) database).

2.2.4. Frequency distribution of amino acids and clustering

The frequency distribution of each amino acid present in ORF10 sequence was determined using standard bioinformatics routine in Matlab [47]. For each ORF10 sequence, a twenty-dimensional frequency-vector considering the frequency of standard twenty amino acids can be obtained. The distance (Euclidean metric) between any two pairs of frequency vectors was calculated for each pair of ORF10 sequences. By having the distance matrix, a set of clusters was obtained applying the well-known *K-means* clustering method using the standard routine in *Matlab-2021a* [48,49].

3. Results

3.1. Continent-wise unique ORF10 variants and mutations

Continent-wise unique variations of ORF10 proteins and their mutations with predicted effects are presented in the following subsections.

3.1.1. ORF10 variants and transmembrane topology of SARS-CoV-2 ORF10

Continent-wise distribution of the total 140 unique ORF10 variants are presented in Table 1. The highest frequency of unique ORF10 variants was found in North America among the total of 140 SARS-CoV-2 ORF10 variants.

Table 1: Unique mutations in the unique ORF10 variants across six continents. # denotes "number of"

	Africa	Asia	Europe	North America	Oceania	South America
# of total ORF10	1183	3393	1033	186572	10290	497
# of unique ORF10	15	21	7	132	15	6
% of unique ORF10	1.27	0.62	0.68	0.07	0.15	1.21
# of residue positions of mutations in ORF10	12	15	6	37	12	5
% of total number of residues	31.6	39.5	15.8	97.4	31.6	13.2

Continent-wise, a decreasing order percentage of unique ORF10 variants was Africa>South America>Europe>Asia>Oceania>North America (Table 1).

In addition, the transmembrane topology and signal peptides for the wild-type SARS-CoV-2 ORF10 (YP_009725255) protein was predicted using the [Phobius](#) webserver. The associated posterior probability for the topology prediction is presented in Figure 2.

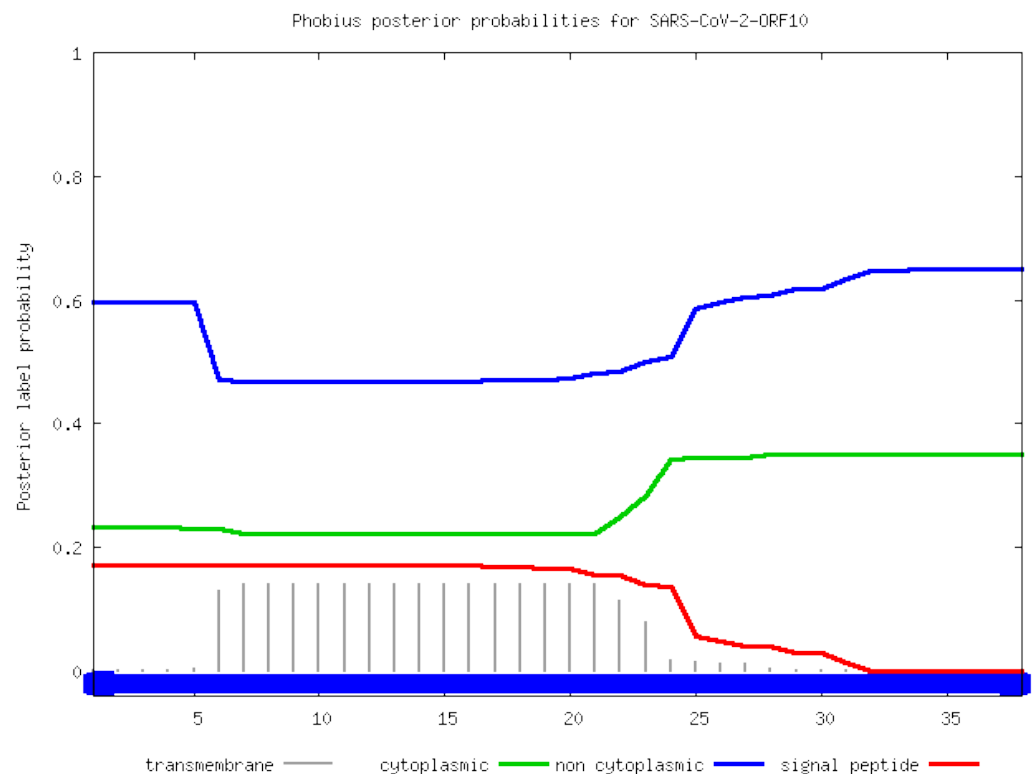


Figure 2. Posterior probability plot for prediction of transmembrane topology and signal peptides for the wild-type SARS-CoV-2 ORF10.

From the posterior probability distributions, it was observed that the SARS-CoV-2 ORF10 protein was entirely non-cytoplasmic (Figure 2).

3.1.2. Unique ORF10 variants and their single and co-occurring mutations

The amino acid residue positions of SARS-CoV-2 ORF10 single-mutations in each continent are listed in Table 2.

Table 2: Continent-wise amino acid positions of single-mutations in SARS-CoV-2 ORF10 unique variants

Continent-wise residue positions of single-mutations					
Africa	Asia	Europe	North America	Oceania	South America
3	6	3	1-38 (except 18th)	3	3
5	7	6		4	19
7	8	13		8	23
10	9	23		10	28
23	10	30		17	35
24	13	31		22	
28	14			23	
30	22			24	
31	23			28	
33	24			30	
37	30			37	
38	31			38	
	33				
	35				
	37				

It was noticed that except for the residue in position 18, all amino acids at each position from 1 to 38 possessed point-missense mutations (Table 2). The common residue was S23, where two mutations S23F and S23P were found across all six continents. A total of eight common residue positions with mutations were observed in Asia and Africa, whereas several common residual mutation positions were associated with Asia and Oceania.

The predicted effects of pathogenicity for each mutation are listed in Table 3. Our data revealed that among a total of 128 mutations, 72 were neutral, and 53 deleterious with regards to pathogenicity.

Table 3: The predicted effect of ORF10 single mutations on pathogenicity

Predicted effect of mutations							
Mutation	Percentage of Pathogenicity		Predcited Type	Mutation	Percentage of Pathogenicity		Predcited Type
	PredictSNP	PhD-SNP			PredictSNP	PhD-SNP	
V33A	83%	66%	Neutral	N22H	83%	51%	Neutral
V33D	87%	77%	Deleterious	N22I	87%	61%	Deleterious
V33F	83%	45%	Neutral	N22K	87%	77%	Deleterious
V33I	83%	78%	Neutral	N22S	83%	68%	Neutral
D31G	87%	73%	Deleterious	N22T	87%	61%	Deleterious
D31H	87%	58%	Deleterious	N22Y	87%	73%	Deleterious
D31N	87%	68%	Deleterious	I4L	83%	78%	Neutral
D31V	87%	82%	Deleterious	I4T	83%	55%	Neutral
D31Y	87%	86%	Deleterious	I4V	83%	83%	Neutral
T38A	83%	78%	Neutral	R20G	87%	58%	Deleterious
T38I	83%	55%	Neutral	R20I	83%	68%	Neutral
L37F	83%	83%	Neutral	R20K	83%	68%	Neutral
L37H	83%	58%	Neutral	R20T	83%	51%	Neutral
L37I	83%	83%	Neutral	M21I	83%	51%	Neutral
L37P	87%	61%	Deleterious	M21K	87%	82%	Deleterious
L37R	87%	61%	Deleterious	M21L	87%	58%	Deleterious
Y3C	83%	58%	Neutral	M21T	87%	73%	Deleterious
Y3H	83%	72%	Neutral	M21V	83%	58%	Neutral
N5D	83%	66%	Neutral	N25D	83%	45%	Neutral
N5K	83%	55%	Neutral	N25G	87%	59%	Deleterious
N5S	83%	66%	Neutral	N25S	83%	55%	Neutral
N5Y	87%	68%	Deleterious	Y26C	87%	68%	Deleterious
R24C	87%	61%	Deleterious	Y26F	83%	68%	Neutral
R24H	83%	51%	Neutral	Y26H	83%	58%	Neutral
R24L	87%	61%	Deleterious	I27K	87%	86%	Deleterious
R24S	83%	58%	Neutral	I27L	83%	51%	Neutral
S23F	87%	58%	Deleterious	I27M	87%	58%	Deleterious
S23P	83%	51%	Neutral	I27R	87%	86%	Deleterious
P10L	87%	86%	Deleterious	I27T	87%	77%	Deleterious
P10Q	7%	73%	Deleterious	I27V	83%	66%	Neutral
P10S	83%	51%	Neutral	D31G	87%	73%	Deleterious
P10T	87%	68%	Deleterious	D31H	87%	58%	Deleterious
F7C	87%	59%	Deleterious	D31N	87%	68%	Deleterious
F7L	83%	55%	Neutral	D31V	87%	82%	Deleterious
F7S	83%	55%	Neutral	D31Y	87%	86%	Deleterious
V6A	83%	78%	Neutral	M1G	83%	72%	Neutral
V6F	83%	55%	Neutral	G2D	83%	68%	Neutral
V6I	83%	78%	Neutral	G2L	83%	58%	Neutral
Y14C	87%	68%	Deleterious	F11L	87%	73%	Deleterious
Y14F	83%	72%	Neutral	F11S	87%	61%	Deleterious
Y14H	83%	55%	Neutral	T12A	83%	58%	Neutral
A28P	87%	73%	Deleterious	T12M	87%	73%	Deleterious
A28S	83%	51%	Neutral	S15C	83%	51%	Neutral
A28V	83%	55%	Neutral	S15G	83%	58%	Neutral
V30A	83%	55%	Neutral	L16P	87%	86%	Deleterious
V30I	83%	78%	Neutral	L17F	83%	72%	Neutral
V30L	83%	51%	Neutral	L17P	87%	86%	Deleterious
F35C	87%	59%	Deleterious	C19F	87%	59%	Deleterious
F35S	83%	51%	Neutral	S23F	87%	58%	Deleterious
A8D	87%	77%	Deleterious	S23P	83%	51%	Neutral
A8G	83%	68%	Neutral	Q29H	83%	55%	Neutral
A8P	87%	73%	Deleterious	Q29L	87%	61%	Deleterious
A8S	83%	55%	Neutral	Q29R	87%	58%	Deleterious
A8V	83%	58%	Neutral	V32A	83%	66%	Neutral
F9L	87%	61%	Deleterious	V32I	83%	78%	Neutral
F9S	87%	58%	Deleterious	V32L	83%	55%	Neutral
F9Y	83%	51%	Neutral	N34D	83%	66%	Neutral
I13L	83%	55%	Neutral	N34Y	87%	73%	Deleterious
I13M	83%	51%	Neutral	N36K	83%	51%	Neutral
I13T	87%	61%	Deleterious	N36S	83%	72%	Neutral
I13V	83%	68%	Neutral	T38A	83%	78%	Neutral
N22D	83%	58%	Neutral	T38I	83%	55%	Neutral
N22F	83%	55%	Neutral				

Among a total of 37 residue positions with single mutations, the residue position 1, 2, 11, 12, 15, 16, 20, 21, 25, 26, 27, 29, 32, 34, and 36 were unique in North America. Among all these residue positions, only the mutations at positions 11 and 16 in ORF10 variants in North America were predicted to be deleterious.

Furthermore, co-mutations in the SARS-CoV-2 ORF10 in some geo-locations were listed in Table 4. The highest number of simultaneous mutations (at 14 positions among the total of 38 amino acid residues) in ORF10 was noticed in a SARS-CoV-2 variant from

Russia on March 21, 2020. Interestingly, no report of this kind of co-occurrent mutation globally, hence denoted by localized index 1.

Table 4: Co-occurrent mutations in the ORF10 variants in various geo-locations

Mutations	Frequency	Date first collected	Localization across countries
USA			
M1K;G2A;Y3D;I4G;N5L;V6Y;F7K;A8R	9	16-01-2021	0.75
P10S;L37F	9	30-11-2020	1
P10L;R24C	2	17-03-2021	1
R24L;A28V	2	01-04-2021	1
C19F;A28V	1	24-03-2021	1
I4V;N5D	1	07-04-2021	1
L37F;T38I	1	09-12-2020	0.5
M1G;G2L	1	18-03-2020	0.4
M1Q;Y3R;I4W;N5A;V6I;A8T	1	23-02-2021	0.5
M1R;G2W;Y3A	1	30-03-2021	0.5
M1R;G2W;Y3A;N5T;V6F	1	22-07-2020	1
M21I;R24C	1	06-04-2021	1
N36C;T38I	1	08-03-2021	1
S23F;I27K	1	02-01-2021	1
UK			
M1G;G2L	2	01-05-2020	0.4
Q29I;V30F;D31N;V32C;N34T;F35L;N36G;L37R	1	29-04-2020	1
R20L;A28V	1	21-05-2020	1
T12M;V30L	1	18-01-2021	0.5
INDIA			
A8L;F9I	1	09-03-2021	1
M1Q;Y3R;I4W;N5A;V6I;A8T	1	11-02-2021	0.5
P10S;F11V	1	23-04-2021	1
SOUTH AFRICA			
R24C;V30L	1	06-01-2021	0.4
SPAIN			
V30L;T38I	20	20-01-2021	0.95
P10S;V30L	13	20-01-2021	0.39
V30L;D31N	8	12-02-2021	0.8
V30L;L37F	4	18-01-2021	0.54
M1L;G2R;Y3P;I4K;N5L;V6M;F7Q	3	11-03-2020	1
R24C;V30L	2	01-03-2021	0.4
S23F;V30L	1	20-01-2021	0.38
GERMANY			
P10S;V30L	15	24-02-2021	0.39
V30L;D31H	4	11-03-2021	1
L16P;V30L	3	16-03-2021	1
P10S;S23F	3	28-03-2021	1
F7V;A8T	1	10-04-2021	1
L17I;R20I;V30L	1	2021-01	1
L37F;T38I	1	06-04-2021	0.5
S23F;V30L	1	2020-12	0.38
T12M;V30L	1	24-03-2021	0.5
GREECE			
I13M;V30L	1	26-02-2021	1
MEXICO			
M1K;G2A;Y3D;I4G;N5L;V6Y;F7K;A8R	1	09-04-2021	0.75
P10S;F35S	1	21-03-2021	1
RUSSIA			
R20T;M21R;S23P;R24A;Y26N;I27A;Q29I;D31L;V32Q;V33L;N34P;F35Q;N36G;L37T	1	21-03-2020	1

Several co-mutations in SARS-CoV-2 ORF10 variants have been reported in the USA, UK, India, South Africa, Spain, Germany, Greece, Mexico, and Russia. The most co-occurrent mutations (V30L, T38I, first reported on January 20, 2021) were found in 20 infected patients in Spain. The US first detected and reported co-occurrent mutations M1K, G2A, Y3D, I4G, N5L, V6Y, F7K, and A8R on January 16, 2021 from 9 infected patients, and later on in April 9, 2021 Germany reported its first case from one patient (Table 4). It is worth knowing that no single mutation at the residue positions 1 and 2 have been reported, and none of the co-mutations among M1K, G2A, Y3D, I4G, N5L, V6Y, F7K, and A8R were reported as a single mutation in the ORF10 variants. Double co-mutations P10S; V30L were reported on January 20, 2021 in Spain on January 20, 2021, and somewhere late February 24, 2021, in Germany. Surprisingly, the pathogenetic effects were predicted to be neutral (Table 3). It was noticed that V30L, one of the most common mutations in ORF10, was co-occurring with most of the other co-mutations. It appears that co-mutations in the SARS-CoV-2 ORF10 variants is an emerging trend.

3.2. Intrinsic disorder regions of SARS-CoV-2 ORF10 variants

The per-residue disorder profiles for 138 unique SARS-CoV-2 ORF10 variants (variants are too short for disorder analysis) and 35 co-occurring mutations are presented in Figure 3. Figure 3A shows that the intrinsic disorder predisposition of ORF10 is noticeably affected by single mutations. Mutations present the largest effects at the N- and C-terminal regions, where the disorder predisposition can vary from 0.4 to 0.8 and from 0.55 to 0.9 for the N- and C-termini, respectively, and for a region centered at residue 25, where the disorder score can change from 0.059 to 0.27. Figure 3B compares the outputs of PONDR-VSL2 for 138 unique ORF10 variants with the results generated for these proteins by one of the most conservative predictors of intrinsic disorder, and IUPred_short [50].

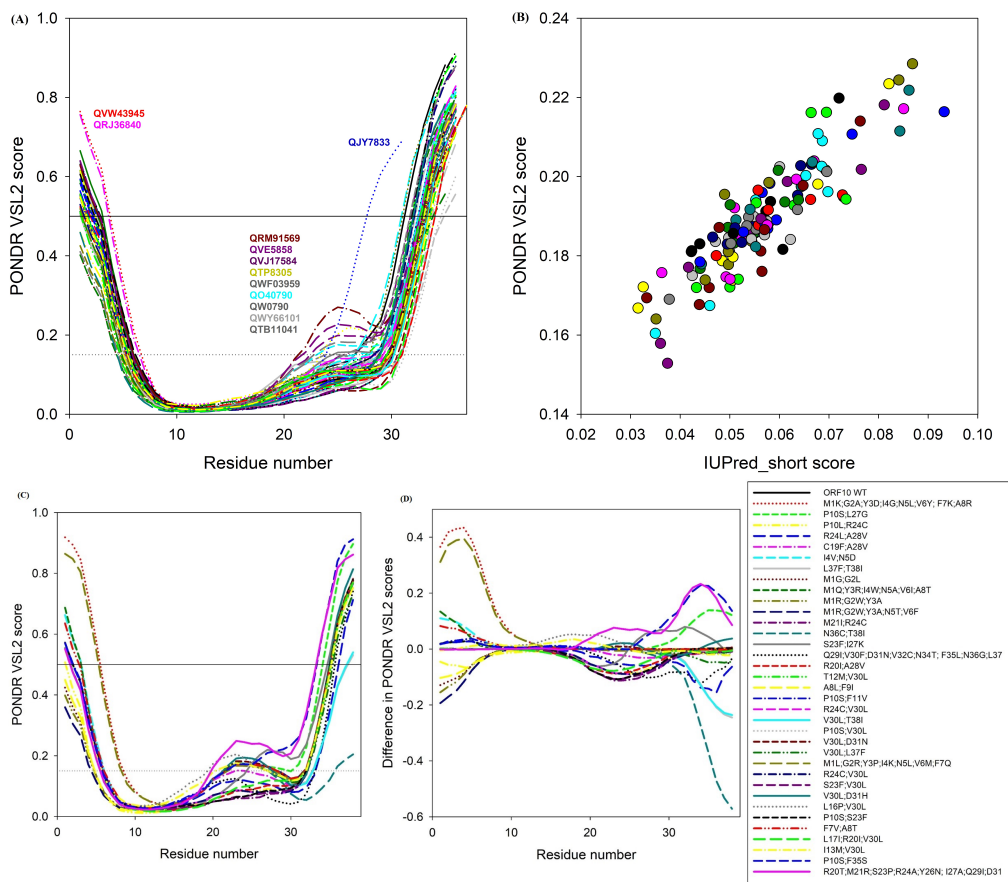


Figure 3. (A): One shows per-residue disorder profiles for 138 SARS-CoV-2 ORF10 variants, (B): Correlation between the outputs of two commonly used disorder predictors, (C) shows how disorder propensity is changed in ORF10 variants with co-occurring mutations, and (D) difference spectra calculated by subtracting off the per-residue disorder scores of wild type ORF10 from the per-residue disorder scores of corresponding mutants.

Here, we investigate the effects of point mutations on the overall disorder score for the entire protein. This analysis showed that although the disorder score values generated by IUPred_short are noticeably smaller than the corresponding PONDR-VSL2 data, the scales of changes introduced by single mutations in intrinsic disorder predisposition of ORF10 as evaluated by IUPred (from 0.03 to 0.095) are comparable to that generated by PONDR VSL2 (from 0.15 to 0.23). Furthermore, both predictors mostly agree on the direction of the changes introduced in the global disorder propensity of this protein by single mutations. Figure 3C shows how disorder propensity is changed in ORF10

variants with co-occurring mutations. Noticeably, the effects of co-occurring mutants are qualitatively similar to those of single mutations, as most variability is observed at the disorder predisposition of N- and C-terminal regions and a region centered at residue 25. However, in the case of co-occurring mutants, scales of changes at the terminal regions are noticeably larger (the disorder predisposition can vary from 0.35 to 0.92 and from 0.2 to 0.9 for the N- and C-termini, respectively). Finally, to simplify comparison of the effect of co-occurring mutations on the disorder predisposition of ORF10, Figure 3D shows "difference spectra" calculated by subtracting of the per-residue disorder scores of wild type ORF10 from the per-residue disorder scores of corresponding mutants. In this plot, intensities of the resulting "bands" reflect the scale of changes, whereas their sign, reflects the mutation-induced increase or decrease (positive or negative values, respectively) in local disorder propensity. Figure 3D shows that most of the co-occurring mutations increase the disorder predisposition of the N-terminal region of ORF10, whereas many such mutations decrease the disorder propensity of the region centered at residue 25, and most of the mutations do not affect disorder predisposition of the C-terminal region.

3.3. Predicting secondary structures of ORF10 protein variants with co-mutations

The SARS-CoV-2 ORF10 protein was found to interact with few of the human proteins such as CUL2, ELOB, ELOC, MAP7D1, PPT1, RBX1, THTPA, TIMM8B, and ZYG11B [33]. A detailed summary of the tissue and cellular expression patterns of SARS-CoV-2 ORF10 interacting human proteins, based on transcriptomics and antibody-based proteomics, are presented in Figure 4. The ORF10 of SARS-CoV-2 interacts with members of the Cullin ubiquitin ligase *CUL2*^{ZYG11B} complex. Interestingly, among the genes ZYG11B scored the highest with regards to ORF10 interactome, which confirmed a direct interaction between ORF10 and ZYG11B [32].

Gene name ⁱ	Covid-19 bait ⁱ	Tissue specificity ⁱ	Tissue expression ⁱ	Blood specificity ⁱ	Subcellular location (main) ⁱ
CUL2	SARS-CoV2 orf10	Low tissue specificity		Low immune cell specificity	Nucleoplasm
ELOB	SARS-CoV2 orf10	Low tissue specificity		Low immune cell specificity	N/A
ELOC	SARS-CoV2 orf10	Low tissue specificity	N/A	Low immune cell specificity	Vesicles
MAP7D1	SARS-CoV2 orf10	Tissue enhanced (skeletal muscle)		Immune cell enhanced (neutrophil)	Microtubules, Cytosol
PPT1	SARS-CoV2 orf10	Low tissue specificity		Low immune cell specificity	Golgi apparatus, Vesicles
RBX1	SARS-CoV2 orf10	Low tissue specificity	N/A	Low immune cell specificity	Nucleoplasm
THTPA	SARS-CoV2 orf10	Low tissue specificity		Low immune cell specificity	Nucleoplasm
TIMM8B	SARS-CoV2 orf10	Low tissue specificity		Low immune cell specificity	N/A
ZYG11B	SARS-CoV2 orf10	Tissue enhanced (skeletal muscle)		Low immune cell specificity	Golgi apparatus, Intermediate filaments

Figure 4. A snapshot summary of the tissue and cellular expression patterns of SARS-CoV-2 ORF10 interacting human proteins, based on transcriptomics and antibody-based proteomics.

The SARS-CoV-2 ORF10 interacting with human proteins CUL2, ELOB, PPT1, THTPA, and TIMM8B were expressed in various tissues, including the lung. SARS-CoV-2 ORF10 contains an α -helical region (amino acid residue positions 3 to 20), which may interact and form a complex with *CUL2*^{ZYG11B} [32]. It was furthermore reported that by forming a complex with *CUL2*^{ZYG11B}, ORF10 hijacks it for ubiquitination and degradation of restriction factors, or alternatively, may binds to the N-terminal glycine in ORF10 to target it for degradation [32,33].

The secondary structures of nine ORF10 variants including the wild ORF10 (YP_009725255) with several co-mutations were predicted (Figure 5). The α -helix secondary structural amino acid (aa) residue position 3 to 20 were invariant for the six ORF variants with co-mutations as mentioned in panel (4) to (9) of Figure 5, although the reliability indices were changed for the α -helix region [32].

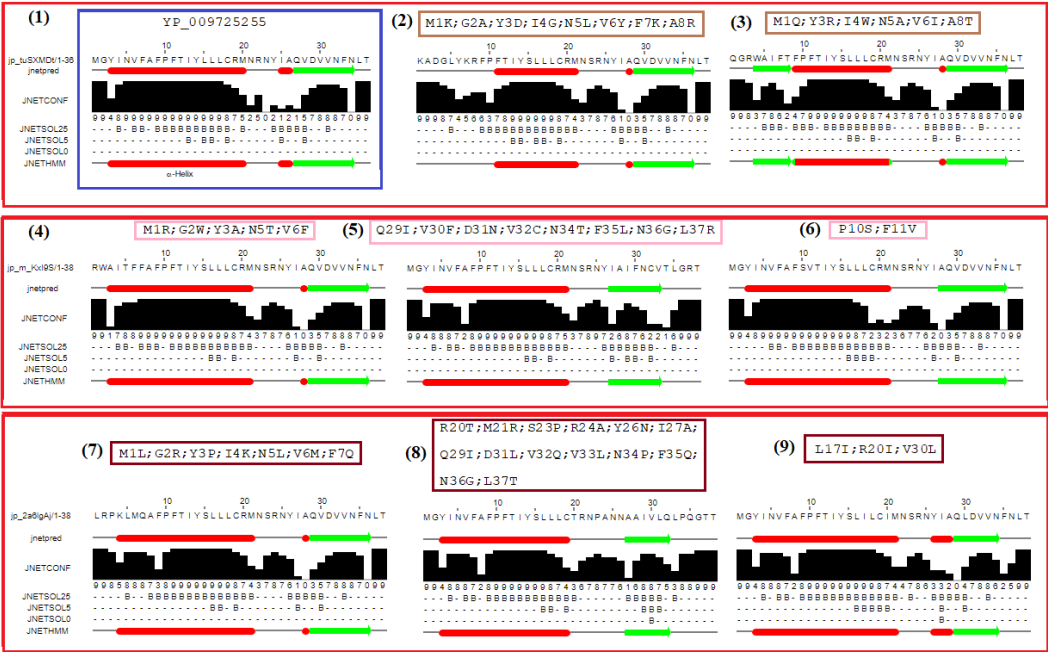


Figure 5. Predicted secondary structure of ORF10 variants with co-mutations.

Furthermore, it was observed that the α –helix amino acid residue positions (aa positions 3-20) got changed to 11-21 and 9-21 due to the co-mutations M1K, G2A, Y3D, I4G, N5L, V6Y, F7K, A8R, and M1Q, Y3R, I4W, N5A, V6I, A8T in two different ORF10 variants, respectively. Therefore, due to these two sets of co-mutations, SARS-CoV-2 ORF10 interactions with the *CUL2*^{ZYG11B} complex might be affected. Due to the amino acid residue positional changes in the secondary structure in the ORF10 variants (other than the reference ORF10 (YP_009725255)), these ORF10 variants with human proteins other than the *CUL2*^{ZYG11B} complex is likely to be affected.

3.4. Amino acid frequency distribution across the unique ORF10 variants and associated clusters

The frequency of each amino acid in the 140 unique ORF10 sequences are presented in Tables 5 and 6. Also, Figure 6(A) shows the amino acid frequency vector of each unique ORF19 sequence. It was noticed that glutamic acid (E) and tryptophan (W) were absent in each ORF10 variant. Among 140 unique ORF10 variants, lysine (K) was present with a single frequency only in seven North American ORF10 variants (QTD22916.1, QLA48060.1, QWY66101.1, QVO40425.1, QTP28305.1, QSL79091.1, and QRM91569.1).

Table 5: Frequency distribution of amino acids over the 140 unique SARS-CoV-2 ORF10 variants (from QKM75696.1 to QWN58574.1)

ORF10	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
QKM75696.1	2	2	4	1	1	1	0	0	0	2	4	0	1	3	1	2	2	0	2	3
QKG88643.1	2	2	5	1	1	1	0	1	0	3	5	0	1	4	1	2	2	0	3	4
QVJ13930.1	2	2	5	2	1	1	0	0	0	3	4	0	2	4	1	2	2	0	3	4
QWE67724.1	2	2	5	1	2	1	0	1	0	3	4	0	2	4	1	2	2	0	2	4
QRJ36840.1	2	2	5	1	1	1	0	1	1	3	4	0	2	4	1	2	2	0	2	4
QQY03084.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QRG22086.1	2	2	4	2	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QTD22916.1	2	2	4	1	1	1	0	1	0	3	4	1	2	4	1	2	2	0	3	4
QUL69971.1	2	2	5	1	1	1	0	1	0	3	4	0	2	2	0	2	2	0	3	3
QWB85197.1	2	2	5	1	1	1	0	1	0	3	4	0	2	5	1	2	2	0	3	3
QWJ83116.1	2	2	5	1	1	1	0	1	0	4	4	0	2	4	1	2	2	0	3	3
QSE09446.1	2	2	5	1	2	1	0	1	0	3	4	0	2	3	1	2	2	0	3	4
QWZ00470.1	2	2	5	1	1	1	0	1	0	3	5	0	2	4	0	2	2	0	3	4
QTP26076.1	3	2	5	1	1	1	0	1	0	3	4	0	2	4	1	2	1	0	3	4
QWK62875.1	2	2	5	1	1	1	0	1	0	3	4	0	3	4	1	2	1	0	3	4
QVG57396.1	2	2	5	1	2	1	0	1	0	3	4	0	2	4	1	2	2	0	2	4
QRA60944.1	2	2	5	1	1	1	0	1	0	3	4	0	2	5	1	2	2	0	2	4
QWE68295.1	2	2	5	1	1	1	0	1	1	3	4	0	2	4	1	2	2	0	2	4
QWT58729.1	2	2	5	1	2	1	0	1	0	3	4	0	2	4	1	1	2	0	3	4
QWM42669.1	2	2	5	1	1	1	0	2	0	3	4	0	2	4	1	1	2	0	3	4
QWS64226.1	2	2	5	1	1	1	0	1	0	3	3	0	2	5	1	2	2	0	3	4
QVJ47956.1	2	1	5	1	1	1	0	2	0	3	4	0	2	4	1	2	2	0	3	4
QVW78101.1	2	1	5	1	1	1	0	1	0	4	4	0	2	4	1	2	2	0	3	4
QLA48060.1	2	1	5	1	1	1	0	1	0	3	4	1	2	4	1	2	2	0	3	4
QVO98764.1	2	2	5	1	1	1	0	1	0	4	4	0	1	4	1	2	2	0	3	4
QWY66101.1	2	2	5	1	1	1	0	1	0	3	4	1	1	4	1	2	2	0	3	4
QWQ05246.1	2	2	5	1	1	1	0	1	0	3	5	0	1	4	1	2	2	0	3	4
QTS35265.1	2	2	4	2	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QVP24786.1	2	2	4	1	1	1	0	1	1	3	4	0	2	4	1	2	2	0	3	4
QVO85840.1	2	2	4	1	1	1	0	1	0	4	4	0	2	4	1	2	2	0	3	4
QVO40425.1	2	2	4	1	1	1	0	1	0	3	4	1	2	4	1	2	2	0	3	4
QWU52456.1	2	2	5	1	1	1	0	1	0	3	4	0	2	5	1	1	2	0	3	4
QTP28305.1	2	2	5	1	1	1	0	1	0	2	4	1	2	5	1	1	2	0	3	4
QWS07290.1	2	2	5	1	1	1	0	1	0	3	4	0	2	4	2	1	2	0	3	4
QWF07009.1	2	1	5	1	2	1	0	1	0	3	3	0	2	5	1	2	2	0	3	4
QWY19801.1	2	1	5	1	2	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QWX30181.1	2	1	5	1	1	1	0	1	1	3	4	0	2	4	1	2	2	0	3	4
QWY95666.1	2	1	5	1	1	1	0	1	0	3	5	0	2	4	1	2	2	0	3	4
QVU00656.1	2	2	4	2	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QTA53643.1	3	2	4	1	1	1	0	2	0	3	4	0	2	4	1	2	2	0	3	3
QUG14309.1	2	2	4	1	1	1	0	2	0	3	4	0	2	4	1	2	2	0	3	4
QWY70751.1	2	2	5	1	2	1	0	1	0	3	4	0	2	4	1	2	2	0	2	4
QSJ35636.1	2	2	5	1	1	1	0	1	0	3	4	0	2	5	1	2	2	0	2	4
QWF03959.1	2	2	5	1	1	1	0	1	1	3	4	0	2	4	1	2	2	0	2	4
BCY15724.1	2	2	3	0	1	0	0	1	0	3	3	0	2	3	1	2	1	0	3	1
QWT72678.1	2	2	5	1	1	0	0	1	1	3	4	0	2	4	1	2	2	0	3	4
QUM42028.1	2	2	5	1	1	0	0	1	0	3	5	0	2	4	1	2	2	0	3	4
QUX49292.1	3	2	5	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	3
QWK69365.1	2	2	5	1	1	1	0	1	0	4	4	0	2	4	1	2	2	0	3	3
QTA74333.1	2	2	5	1	1	1	0	1	0	3	4	0	2	5	1	2	2	0	3	3
QWY54619.1	2	2	5	1	1	1	0	1	0	3	5	0	2	4	1	2	2	0	3	3
QUV26065.1	3	2	5	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	3
QWF05003.1	2	2	5	1	1	1	0	1	0	4	4	0	2	4	1	2	2	0	3	3
QUP00476.1	2	2	5	1	1	1	0	1	0	3	5	0	2	4	1	2	2	0	3	3
QIJ93574.1	3	2	5	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	3
QWN49685.1	2	2	5	2	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	3
QVO91006.1	2	2	5	1	1	1	0	1	0	3	4	0	2	5	1	2	2	0	3	3
QVV18442.1	2	2	5	1	1	1	0	1	0	4	4	0	2	4	1	2	2	0	3	3
QJY78233.1	2	2	3	1	1	1	0	1	0	3	3	0	2	3	1	2	1	0	3	4
QVL90897.1	2	2	4	2	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QOH29638.1	2	2	5	1	2	1	0	1	0	3	4	0	2	3	1	2	2	0	3	4
QWC74916.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QSL79091.1	2	2	4	1	1	1	0	1	0	3	4	1	2	4	1	2	2	0	3	4
BCX23983.1	2	2	5	1	1	1	0	1	0	3	3	0	2	4	1	2	1	0	3	4
QWY94400.1	2	2	5	1	1	1	0	1	0	3	3	0	2	5	1	2	2	0	3	4
QSE25736.1	2	2	5	1	1	1	0	1	1	3	3	0	2	4	1	2	2	0	3	4
QWN58574.1	2	2	5	1	1	1	0	1	0	4	3	0	2	4	1	2	2	0	3	4

Table 6: Frequency distribution of amino acids over the 140 unique SARS-CoV-2 ORF10 variants (from QTI75440.1 to QWY55420.1)

ORF10	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
QTI75440.1	3	2	5	1	1	1	0	1	0	3	4	0	2	4	1	2	1	0	3	4
QWS53172.1	2	2	5	1	1	1	0	1	0	4	4	0	2	4	1	2	1	0	3	4
YP_009725255.1	2	2	5	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QWY95092.1	2	2	5	1	1	1	0	1	0	3	3	0	2	4	2	2	2	0	3	4
QWQ66744.1	2	3	5	1	1	1	0	1	0	3	3	0	2	4	1	2	2	0	3	4
QSW62483.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	3	2	0	3	4
QVU28280.1	2	2	5	1	1	1	0	1	0	3	5	0	2	3	1	2	2	0	3	4
QWW53635.1	2	2	5	1	1	1	0	1	0	3	4	0	2	3	1	3	2	0	3	4
QQX05795.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	4	4
QVE28736.1	2	2	5	0	1	1	0	2	0	3	4	0	2	4	1	2	2	0	3	4
QTC84700.1	2	2	5	0	1	1	0	1	1	3	4	0	2	4	1	2	2	0	3	4
QWN49673.1	2	2	6	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	3
QUA36764.1	2	2	6	0	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QSW42096.1	2	2	5	0	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	5
QWY22460.1	2	2	5	0	1	1	0	1	0	3	4	0	2	4	1	2	2	0	4	4
QUF19963.1	2	3	5	1	1	0	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QTB11041.1	1	2	5	1	1	1	0	1	0	3	4	0	2	4	2	2	2	0	3	4
QWX09518.1	1	2	5	1	1	1	0	1	0	3	4	0	2	4	1	3	2	0	3	4
QWU01215.1	1	2	5	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	5
QRM91569.1	2	2	5	1	1	1	0	1	0	2	4	1	2	4	1	2	2	0	3	4
QRX03618.1	2	2	5	1	1	1	0	1	0	2	5	0	2	4	1	2	2	0	3	4
QSO40790.1	2	2	5	1	1	1	0	1	0	2	4	0	3	4	1	2	2	0	3	4
QVE25858.1	2	3	5	1	1	1	0	1	0	2	4	0	2	4	1	2	2	0	3	4
QVJ17584.1	2	2	5	1	1	1	0	1	0	2	4	0	2	4	1	2	3	0	3	4
QVO95200.1	2	2	5	1	1	1	0	1	0	2	4	0	2	4	1	2	2	0	3	5
QVE30392.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	3	2	0	3	4
QTC19517.1	2	1	5	1	1	1	0	1	0	3	4	0	2	4	1	3	2	0	3	4
QVJ37366.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	3	2	0	3	4
QWQ76823.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	2	3	0	3	4
QVV08801.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	4	4
QUI12106.1	2	2	5	1	1	1	0	1	0	3	4	0	1	4	1	2	3	0	3	4
QWU53472.1	2	2	5	1	1	1	0	1	0	3	4	0	1	4	1	2	2	0	3	5
QTS24551.1	2	1	5	1	1	1	0	1	0	3	4	0	2	4	1	2	3	0	3	4
QNV50343.1	2	2	5	1	0	1	0	1	0	3	4	0	2	5	1	2	2	0	3	4
QUB17908.1	1	2	5	1	0	1	0	1	0	3	4	0	2	5	1	2	2	0	3	5
QVL64016.1	2	2	5	1	1	1	0	1	0	3	3	0	2	4	2	2	2	0	3	4
QWU68360.1	2	2	5	1	1	1	0	1	0	3	3	0	2	4	2	2	2	0	3	4
QTO29824.1	2	2	5	1	1	1	0	1	0	2	5	0	2	4	1	2	2	0	3	4
QWU51246.1	2	2	5	1	1	1	0	1	0	2	4	0	3	4	1	2	2	0	3	4
QVX69392.1	2	2	5	1	1	1	0	1	0	2	4	0	2	4	1	2	3	0	3	4
QWO21857.1	2	2	5	1	1	1	0	1	0	2	4	0	2	4	1	2	2	0	3	5
QTK02152.1	2	2	5	1	1	1	0	1	0	3	5	0	2	3	1	2	2	0	3	4
QWY72735.1	2	2	5	1	1	1	0	1	0	3	4	0	2	3	1	3	2	0	3	4
QVH90751.1	2	2	5	1	1	2	0	1	0	3	4	0	2	4	0	2	2	0	3	4
QVX36355.1	2	1	5	1	1	1	0	1	0	3	5	0	2	4	0	3	2	0	3	4
QWY95104.1	2	2	5	1	1	1	0	1	0	3	3	0	2	5	0	3	2	0	3	4
QWY18545.1	2	2	5	1	1	1	0	1	0	3	4	0	2	4	0	3	2	0	3	4
QVL15727.1	2	2	5	1	1	1	0	1	0	3	4	0	2	4	0	2	3	0	3	4
QTO07027.1	2	2	5	1	1	1	0	1	0	3	4	0	2	3	0	2	2	0	3	4
BCX25240.1	2	2	5	1	1	1	0	1	0	3	5	0	2	3	1	2	2	0	3	4
QWB65585.1	2	2	5	1	1	1	0	1	0	3	4	0	2	3	1	3	2	0	3	4
QTW57386.1	2	2	5	1	1	1	0	1	0	3	4	0	2	3	1	2	2	0	4	4
QWW38212.1	1	2	5	2	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	4
QSE30156.1	1	2	5	1	1	1	0	2	0	3	4	0	2	4	1	2	2	0	3	4
QUF17645.1	1	2	5	1	1	1	0	1	0	3	4	0	2	4	2	2	2	0	3	4
QUF20717.1	1	2	5	1	1	1	0	1	0	3	4	0	2	3	1	2	2	0	3	4
QWB83606.1	1	2	5	1	1	1	0	1	0	3	4	0	2	4	1	3	2	0	3	4
QPF60767.1	1	2	5	1	1	1	0	1	0	3	4	0	2	4	1	2	3	0	3	4
QVR42407.1	1	2	5	1	1	1	0	1	0	3	4	0	2	2	0	2	2	0	3	4
QUA32182.1	1	2	5	1	1	1	0	1	0	3	5	0	2	4	0	2	2	0	3	5
QRG41735.1	1	1	5	1	1	1	0	1	0	3	5	0	2	4	1	2	2	0	3	5
QWT73590.1	1	2	5	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	3	5
QWK62266.1	2	2	5	1	1	1	0	1	0	3	5	0	2	3	1	2	2	0	3	4
QVM67662.1	2	2	5	1	1	1	0	1	0	4	4	0	2	3	1	3	1	0	3	4
QQE14148.1	2	2	5	1	1	1	0	1	0	3	4	0	2	3	1	3	2	0	3	4
QVU00728.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	3	2	0	3	4
QUA79573.1	2	2	4	1	1	1	0	1	0	3	4	0	2	4	1	2	2	0	4	4
QWT65692.1	2	2	5	1	1	1	0	1	0	2	5	0	2	4	1	2	2	0	3	4
QVW43945.1	2	2	5	1	1	1	0	1	0	2	4	0	2	4	1	2	3	0	3	4
QUM21266.1	2	2	4	2	1	1	0	1	0	2	4	0	2	4	1	2	2	0	3	5
QTX06599.1	2	1	5	1	2	1	0	1	0	2	4	0	2	4	1	2	2	0	3	5
QVM14069.1	2	2	5	1	1	1	0	1	0	2	3	0	2	5	1	2	2	0	3	5
QWY55420.1	2	2	5	1	1	1	0	1	0	2	4	0	2	4	1	2	2	0	3	5

It was noticed that asparagine was present in the ORF10 sequences with the highest frequencies ranging from 4 to 6 (Figure 6(A)). Also, the frequencies of L, F, and V in each

ORF10 variant were dominant as compared to that of other amino acids.

For each pair of frequency vectors associated with ORF10 sequences, distances were enumerated (Figure 6(B)).

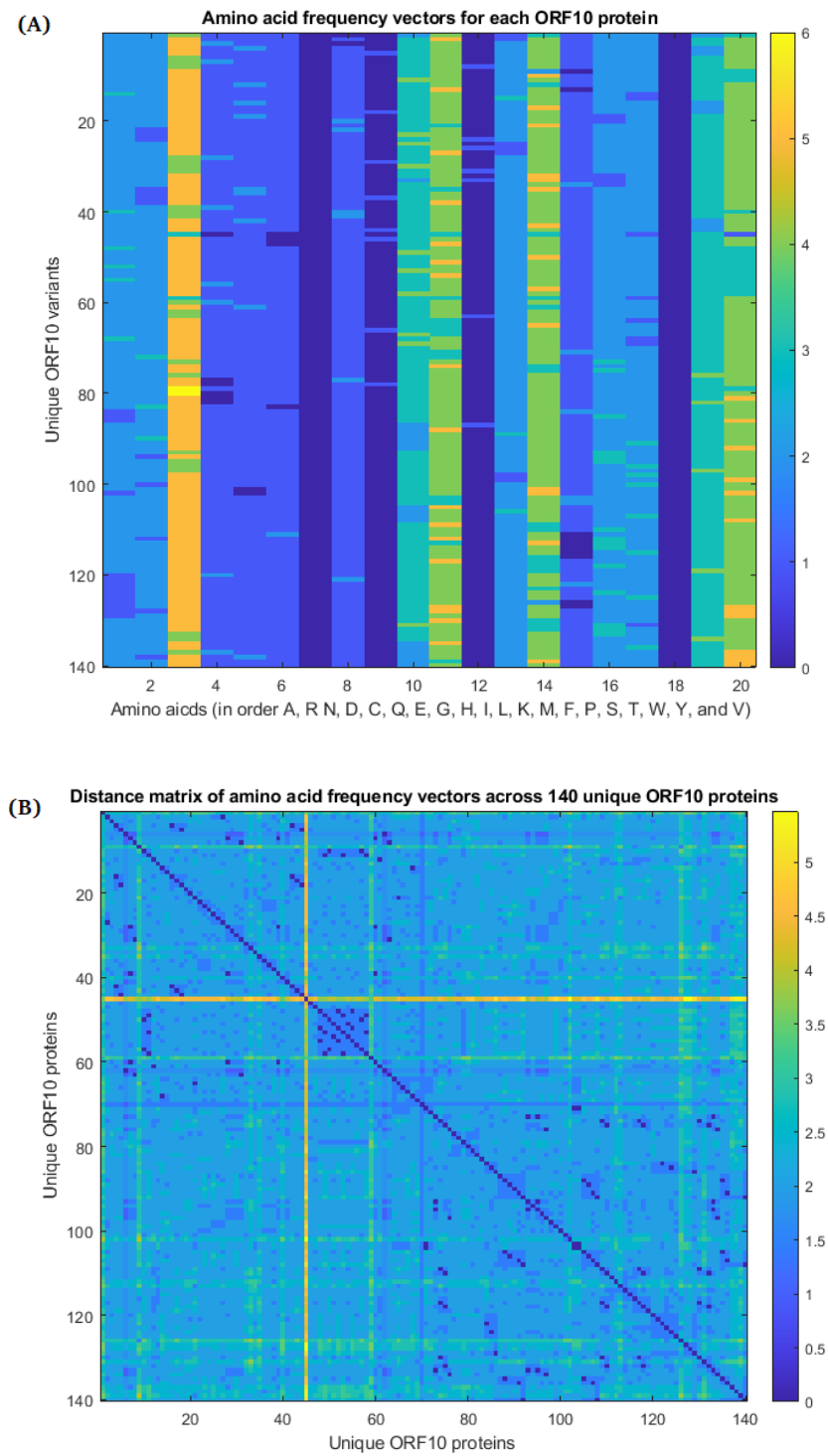


Figure 6. (A): Frequency distribution of amino acid across the unique 140 ORF10 proteins, and (B) Pairwise distance matrix of frequency vectors for each ORF10 protein

The most distant frequency vector was detected for the SARS-CoV-2 ORF10 BCY15724.1 (Japan, collected in November, 2020). Based on frequency-vector similarity (Euclidean distance-wise nearness), unique 140 ORF10 proteins were clustered using the K-means clustering method (Table 7).

Table 7: K-means clustering (clusters: 20) of the unique ORF10 variants based on amino acid frequency distribution

ORF10	Cluster	ORF10	Cluster	ORF10	Cluster	ORF10	Cluster
QKM75696.1	8	QVW78101.1	11	BCY15724.1	17	QWN58574.1	15
QKG88643.1	1	QLA48060.1	1	QWT72678.1	16	QTI75440.1	16
QVJ13930.1	1	QVO98764.1	11	QUM42028.1	1	QWS53172.1	11
QWE67724.1	7	QWY66101.1	1	QUX49292.1	14	YP_009725255.1	5
QRJ36840.1	7	QWQ05246.1	1	QWK69365.1	11	QWY95092.1	15
QQY03084.1	5	QTS35265.1	20	QTA74333.1	3	QWQ66744.1	15
QRG22086.1	20	QVP24786.1	20	QWY54619.1	14	QSW62483.1	5
QTD22916.1	20	QVO85840.1	20	QUV26065.1	14	QVU28280.1	6
QUL69971.1	18	QVO40425.1	20	QWF05003.1	11	QWW53635.1	19
QWB85197.1	3	QWU52456.1	3	QUP00476.1	14	QQX05795.1	20
QWJ83116.1	11	QTP28305.1	10	QTI93574.1	14	QVE28736.1	16
QSE09446.1	6	QWS07290.1	16	QWN49685.1	14	QTC84700.1	16
QWZ00470.1	1	QWF07009.1	10	QVO91006.1	3	QWN49673.1	14
QTP26076.1	16	QWY19801.1	1	QVV18442.1	11	QUA36764.1	16
QWK62875.1	16	QWX30181.1	1	QJY78233.1	8	QSW42096.1	4
QVG57396.1	7	QWY95666.1	1	QVL90897.1	20	QWY22460.1	16
QRA60944.1	3	QVU00656.1	20	QOH29638.1	6	QUF19963.1	16
QWE68295.1	7	QTA53643.1	16	QWC74916.1	5	QTB11041.1	4
QWT58729.1	16	QUG14309.1	20	QSL79091.1	20	QWX09518.1	4
QWM42669.1	16	QWY70751.1	7	BCX23983.1	15	QWU01215.1	4
QWS64226.1	3	QSJ35636.1	3	QWY94400.1	3	QRM91569.1	9
QVJ47956.1	1	QWF03959.1	7	QSE25736.1	15	QRX03618.1	12

ORF10	Cluster	ORF10	Cluster	ORF10	Cluster	ORF10	Cluster
QSO40790.1	9	QVH90751.1	1	QVU00728.1	5	QWU68360.1	15
QVE25858.1	9	QVX36355.1	2	QUA79573.1	20	QTO29824.1	12
QVJ17584.1	9	QWY95104.1	10	QWT65692.1	12	QWU51246.1	9
QVO95200.1	9	QWY18545.1	1	QVW43945.1	9	QVX69392.1	9
QVE30392.1	5	QVL15727.1	1	QUM21266.1	10	QWO21857.1	9
QTC19517.1	1	QTO07027.1	6	QTX06599.1	10	QTK02152.1	6
QVJ37366.1	5	BCX25240.1	6	QVM14069.1	10	QWY72735.1	19
QWQ76823.1	20	QWB65585.1	19	QWY55420.1	9		
QVV08801.1	20	QTW57386.1	6	QVR42407.1	18		
QUI12106.1	1	QWW38212.1	4	QUA32182.1	2		
QWU53472.1	4	QSE30156.1	4	QRG41735.1	2		
QTS24551.1	1	QUF17645.1	4	QWT73590.1	4		
QNV50343.1	3	QUF20717.1	6	QWK62266.1	6		
QUB17908.1	2	QWB83606.1	4	QVM67662.1	13		
QVL64016.1	15	QPF60767.1	4	QQE14148.1	19		

Based on amino acid compositions in each ORF10 sequence, unique ORF10 proteins were clustered into 20 different clusters. Cluster-1 contains the highest number (17) of ORF10 variants, which fall in the US, whereas clusters 13 and 17 contain only one ORF10 variant each. The ORF10 proteins BCY15724 (collected from Japan) and QVM67662 (USA:Florida) belong to cluster 13 and 17, respectively (Table 8).

Table 8: Frequency of ORF10 variants in each cluster

Cluster Number	Frequency	Cluster Number	Frequency	Cluster Number	Frequency
1	17	5	7	19	4
20	14	11	7	12	3
16	13	14	7	8	2
4	11	15	7	18	2
9	10	7	6	13	1
3	9	10	6	17	1
6	9	2	4		

Clearly, twenty different clusters with different frequencies of ORF10 variants showed wide variations of ORF10 sequences based on amino acid compositions.

4. Discussion

In this study, a total of 140 unique SARS-CoV-2 ORF10 sequences were observed among 202968 ORF10 sequences obtained from the NCBI database. Remarkably, most unique ORF10 variants were first reported in North America only (Table 1). Consequently, the highest number of point mutations and co-mutation were found in North America. Continent-wise Africa had the highest percentage (1.27%) of unique variants of ORF10 sequences from Africa. Furthermore, it was observed that all amino acids of ORF10 except the residue 18, possessed single-missense mutations (Table 2). The unique ORF10 variants from Asia, Africa, and Oceania had several point-mutations at common residue positions (Table 2). We noticed that in North America, ORF10 mutations only at the amino acid residue positions 11 (F11S, F11L) and 16 (L16P) were deleterious, as reported from North America. A significant percentage of non-synonymous mutations, 53 among the total of 128 mutations were deleterious, which may alter the intensity of the interactome between ORF10 and host proteins. The 22 unique co-mutations were observed in different geo-locations Table 4. It was noticed that co-mutations in ORF10 variants became an emerging trend, which varies slightly in 2020, and inevitably in the future ORF10 variants with co-mutations might be transmitted to other geo-locations while restriction on international travel is lifted. Furthermore, it was observed that the effects of co-occurring mutants are qualitatively similar to those of single mutations, as most variability is at the disorder predisposition of N- and C-terminal regions and a region centered at residue 25. However, in co-occurring mutants, scales of changes at the terminal regions are noticeably larger. Most of the co-occurring mutations increase the disorder predisposition of N-terminal region of ORF10, whereas many such mutations decrease the disorder propensity of the region centered at residue 25. Consequently, these mutations may affect secondary structure and associated functions of the ORF10 variants. In addition, amino acid residue positions of the secondary structures (especially the N-terminal α -helix regions) of several ORF10 variants changed, and therefore interaction of those ORF10 variants with the human protein *CUL2*^{ZYG11B} complex are most likely affected.

5. Conclusions

Unique variations in SARS-CoV-2 ORF10 proteins are an emerging trend across different continents due to the appearance of various single point and co-occurring mutations. The highest percentage of unique ORF10 variants were found in Africa, though the frequency of unique ORF10 variants was significantly dominating with a total number of unique variants available. It was further observed that the growth rate of emerging non-synonymous mutations in ORF10 proteins is increasing non-linearly, which is certainly alarming with regards to stability or instability of emerging SARS-CoV-2 variants. Due to significant deleterious mutations, expression of the SARS-CoV-2 ORF10 proteins might get altered a phenomenon, which will affect functional virus-host protein-protein interactions. Also, it was reported that secondary structure

α –helix regions of the ORF10 variants was changed due to co-mutations. Consequently, virulence/pathogenicity of SARS-CoV-2 may get influenced directly or indirectly and therefore, continuous surveillance of mutations and their associated effects is necessary.

Author Contributions: SSH and VU designed research and performed research. SSH, VU wrote the primary draft of the article. KL, ASA, PA, AAAA, AL, RK, TMAEA, PPC, GKA, SS, MT, GC, KT, DB, and GP edited the manuscript. EMR and PB proofread the final article. All authors read the final article, and approved for submission.

Funding: This research received no external funding.

Acknowledgments: We are grateful to Agnel Praveen Joseph of the Science and Technology Facilities Council, UKRI, UK for his kind help in getting several information related to the webserver CoVal.

Conflicts of Interest: Authors declare no competing interest to declare.

References

1. Worldometer. COVID Live Update. <https://www.worldometers.info/coronavirus/>, 2010 (accessed July 16, 2021).
2. Johns-Hopkin-University. Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html>, 2010 (accessed July 16, 2021).
3. Laha, S.; Chakraborty, J.; Das, S.; Manna, S.K.; Biswas, S.; Chatterjee, R. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infection, Genetics and Evolution* **2020**, *85*, 104445.
4. Kim, J.S.; Jang, J.H.; Kim, J.M.; Chung, Y.S.; Yoo, C.K.; Han, M.G. Genome-wide identification and characterization of point mutations in the SARS-CoV-2 genome. *Osong Public Health and Research Perspectives* **2020**, *11*, 101.
5. Weber, S.; Ramirez, C.M.; Weiser, B.; Burger, H.; Doerfler, W. SARS-CoV-2 worldwide replication drives rapid rise and selection of mutations across the viral genome: a time-course study–potential challenge for vaccines and therapies. *EMBO Molecular Medicine* **2021**, p. e14062.
6. Krause, P.R.; Fleming, T.R.; Longini, I.M.; Peto, R.; Briand, S.; Heymann, D.L.; Beral, V.; Snape, M.D.; Rees, H.; Roper, A.M.; others. SARS-CoV-2 Variants and Vaccines. *New England Journal of Medicine* **2021**.
7. Touret, F.; Luciani, L.; Baronti, C.; Cochin, M.; Driouich, J.S.; Gilles, M.; Thirion, L.; Nougairède, A.; de Lamballerie, X. Replicative Fitness of a SARS-CoV-2 201/501Y. V1 Variant from Lineage B. 1.1. 7 in Human Reconstituted Bronchial Epithelium. *Mbio* **2021**, *12*, e00850–21.
8. Maurano, M.T.; Ramaswami, S.; Zappile, P.; Dimartino, D.; Boytard, L.; Ribeiro-dos Santos, A.M.; Vulpescu, N.A.; Westby, G.; Shen, G.; Feng, X.; others. Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York City region. *Genome research* **2020**, *30*, 1781–1788.
9. Bi, C.; Ramos-Mandujano, G.; Tian, Y.; Hala, S.; Xu, J.; Mfarrej, S.; Esteban, C.R.; Delicado, E.N.; Alofi, F.S.; Khogeer, A.; others. Simultaneous detection and mutation surveillance of SARS-CoV-2 and multiple respiratory viruses by rapid field-deployable sequencing. *Med* **2021**.
10. Satarker, S.; Nampoothiri, M. Structural proteins in severe acute respiratory syndrome coronavirus-2. *Archives of medical research* **2020**, *51*, 482–491.
11. Wang, D.; Jiang, A.; Feng, J.; Li, G.; Guo, D.; Sajid, M.; Wu, K.; Zhang, Q.; Ponty, Y.; Will, S.; others. The SARS-CoV-2 subgenome landscape and its novel regulatory features. *Molecular cell* **2021**, *81*, 2135–2147.
12. Kasibhatla, S.M.; Kinikar, M.; Limaye, S.; Kale, M.M.; Kulkarni-Kale, U. Understanding evolution of SARS-CoV-2: a perspective from analysis of genetic diversity of RdRp gene. *Journal of medical virology* **2020**, *92*, 1932–1937.
13. Michel, C.J.; Mayer, C.; Poch, O.; Thompson, J.D. Characterization of accessory genes in coronavirus genomes. *Virology Journal* **2020**, *17*, 1–13.
14. Silva, S.J.R.d.; Silva, C.T.A.d.; Mendes, R.P.G.; Pena, L.; others. Role of nonstructural proteins in the pathogenesis of SARS-CoV-2 **2020**.
15. Hillen, H.S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of replicating SARS-CoV-2 polymerase. *Nature* **2020**, *584*, 154–156.
16. Clark, L.K.; Green, T.J.; Petit, C.M. Structure of nonstructural protein 1 from SARS-CoV-2. *Journal of Virology* **2021**, *95*, e02019–20.
17. Shannon, A.; Le, N.T.T.; Selisko, B.; Eydoux, C.; Alvarez, K.; Guillemot, J.C.; Decroly, E.; Peersen, O.; Ferron, F.; Canard, B. Remdesivir and SARS-CoV-2: Structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. *Antiviral research* **2020**, *178*, 104793.
18. Duffy, S. Why are RNA virus mutation rates so damn high? *PLoS biology* **2018**, *16*, e3000003.
19. Elena, S.F.; Sanjuán, R. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *Journal of virology* **2005**, *79*, 11555–11558.
20. Yan, L.; Yang, Y.; Li, M.; Zhang, Y.; Zheng, L.; Ge, J.; Huang, Y.C.; Liu, Z.; Wang, T.; Gao, S.; others. Coupling of N7-methyltransferase and 3′-5′ exonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell* **2021**, *184*, 3474–3485.

21. Chen, J.; Wang, R.; Wang, M.; Wei, G.W. Mutations strengthened SARS-CoV-2 infectivity. *Journal of molecular biology* **2020**, *432*, 5212–5226.
22. Eskier, D.; Karakülah, G.; Suner, A.; Oktay, Y. RdRp mutations are associated with SARS-CoV-2 genome evolution. *PeerJ* **2020**, *8*, e9587.
23. Gupta, A.M.; Chakrabarti, J.; Mandal, S. Non-synonymous mutations of SARS-CoV-2 leads epitope loss and segregates its variants. *Microbes and infection* **2020**, *22*, 598–607.
24. Hassan, S.S.; Choudhury, P.P.; Roy, B. SARS-CoV2 envelope protein: non-synonymous mutations and its consequences. *Genomics* **2020**, *112*, 3890–3892.
25. Issa, E.; Merhi, G.; Panossian, B.; Salloum, T.; Tokajian, S. SARS-CoV-2 and ORF3a: nonsynonymous mutations, functional domains, and viral pathogenesis. *Msystems* **2020**, *5*, e00266–20.
26. Seyran, M.; Pizzol, D.; Adadi, P.; El-Aziz, T.M.; Hassan, S.S.; Soares, A.; Kandimalla, R.; Lundstrom, K.; Tambuwala, M.; Aljabali, A.A.; others. Questions concerning the proximal origin of SARS-CoV-2. *Journal of Medical Virology* **2021**, *93*, 1204.
27. Narayanan, K.; Huang, C.; Makino, S. SARS coronavirus accessory proteins. *Virus research* **2008**, *133*, 113–121.
28. Hassan, S.S.; Attrish, D.; Ghosh, S.; Choudhury, P.P.; Uversky, V.N.; Aljabali, A.A.; Lundstrom, K.; Uhal, B.D.; Rezaei, N.; Seyran, M.; others. Notable sequence homology of the ORF10 protein introspects the architecture of SARS-COV-2. *International Journal of Biological Macromolecules* **2021**, *181*, 801–809.
29. Schuster, N.A. Characterization and structural prediction of the putative ORF10 protein in SARS-CoV-2. *bioRxiv* **2021**, pp. 2020–10.
30. Pancer, K.; Milewska, A.; Owczarek, K.; Dabrowska, A.; Kowalski, M.; Labaj, P.P.; Branicki, W.; Sanak, M.; Pyrc, K. The SARS-CoV-2 ORF10 is not essential in vitro or in vivo in humans. *PLoS Pathogens* **2020**, *16*, e1008959.
31. Altincekic, N.; Korn, S.M.; Qureshi, N.S.; Dujardin, M.; Ninot-Pedrosa, M.; Abele, R.; Abi Saad, M.J.; Alfano, C.; Almeida, F.C.; Alshamleh, I.; others. Large-scale recombinant production of the SARS-CoV-2 proteome for high-throughput and structural biology applications. *Frontiers in molecular biosciences* **2021**, *8*, 89.
32. Gordon, D.E.; Jang, G.M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K.M.; O'Meara, M.J.; Rezelj, V.V.; Guo, J.Z.; Swaney, D.L.; others. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583*, 459–468.
33. Mena, E.L.; Donahue, C.J.; Vaites, L.P.; Li, J.; Rona, G.; O'Leary, C.; Lignitto, L.; Miwatani-Minter, B.; Paulo, J.A.; Dhabaria, A.; others. ORF10–Cullin-2–ZYG11B complex is not required for SARS-CoV-2 infection. *Proceedings of the National Academy of Sciences* **2021**, *118*.
34. Li, J.; Guo, M.; Tian, X.; Wang, X.; Yang, X.; Wu, P.; Liu, C.; Xiao, Z.; Qu, Y.; Yin, Y.; others. Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med* **2021**, *2*, 99–112.
35. Yang, D.M.; Lin, F.C.; Tsai, P.H.; Chien, Y.; Wang, M.L.; Yang, Y.P.; Chang, T.J. Pandemic analysis of infection and death correlated with genomic open reading frame 10 mutation in severe acute respiratory syndrome coronavirus 2 victims. *Journal of the Chinese Medical Association* **2021**, *84*, 478–484.
36. Cagliani, R.; Forni, D.; Clerici, M.; Sironi, M. Coding potential and sequence conservation of SARS-CoV-2 and related animal viruses. *Infection, Genetics and Evolution* **2020**, *83*, 104353.
37. Käll, L.; Krogh, A.; Sonnhammer, E.L. A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology* **2004**, *338*, 1027–1036.
38. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.; Potter, S.C.; Finn, R.D.; others. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research* **2019**, *47*, W636–W641.
39. Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G.J. JPred4: a protein secondary structure prediction server. *Nucleic acids research* **2015**, *43*, W389–W394.
40. Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Dunker, A.K. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins: Structure, Function, and Bioinformatics* **2005**, *61*, 176–182.
41. Meng, F.; Uversky, V.N.; Kurgan, L. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cellular and Molecular Life Sciences* **2017**, *74*, 3069–3090.
42. Peng, Z.L.; Kurgan, L. Comprehensive comparative assessment of in-silico predictors of disordered regions. *Current Protein and Peptide Science* **2012**, *13*, 6–18.
43. Fan, X.; Kurgan, L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *Journal of Biomolecular Structure and Dynamics* **2014**, *32*, 448–464.
44. Pickett, B.E.; Sadat, E.L.; Zhang, Y.; Noronha, J.M.; Squires, R.B.; Hunt, V.; Liu, M.; Kumar, S.; Zaremba, S.; Gu, Z.; others. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic acids research* **2012**, *40*, D593–D598.
45. Bendl, J.; Stourac, J.; Salanda, O.; Pavelka, A.; Wieben, E.D.; Zendulka, J.; Brezovsky, J.; Damborsky, J. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS computational biology* **2014**, *10*, e1003440.
46. Capriotti, E.; Fariselli, P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic acids research* **2017**, *45*, W247–W252.
47. Henson, R.; Cetto, L. The MATLAB bioinformatics toolbox. *Encyclopedia of genetics, genomics, proteomics and bioinformatics* **2004**.
48. Hassan, S.S.; Aljabali, A.A.; Panda, P.K.; Ghosh, S.; Attrish, D.; Choudhury, P.P.; Seyran, M.; Pizzol, D.; Adadi, P.; Abd El-Aziz, T.M.; others. A unique view of SARS-CoV-2 through the lens of ORF8 protein. *Computers in biology and medicine* **2021**, *133*, 104380.

-
49. Hassan, S.S.; Choudhury, P.P.; Roy, B.; Jana, S.S. Missense mutations in SARS-CoV2 genomes from Indian patients. *Genomics* **2020**, *112*, 4622–4627.
 50. Mészáros, B.; Erdős, G.; Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic acids research* **2018**, *46*, W329–W337.