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.

Archives of Biochemistry and Biophysics 717 (2022) 109124

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

The importance of accessory protein variants in the pathogenicity of SARS-CoV-2

Sk. Sarif Hassan [a,**], Pabitra Pal Choudhury [b], Guy W. Dayhoff II [c], Alaa A.A. Aljabali [d], Bruce D. Uhal [e], Kenneth Lundstrom [f,***], Nima Rezaei [g,af], Damiano Pizzol [h], Parise Adadi [i], Amos Lal [j], Antonio Soares [k], Tarek Mohamed Abd El-Aziz [k,l], Adam M. Brufsky [m], Gajendra Kumar Azad [n], Samendra P. Sherchan [o], Wagner Baetas-da-Cruz [p], Kazuo Takayama [q], Ãngel Serrano-Aroca [r], Gaurav Chauhan [s], Giorgio Palu [t], Yogendra Kumar Mishra [u], Debmalya Barh [v,w], Raner Jośe Santana Silva [x], Bruno Silva Andrade [y], Vasco Azevedo [w], Aristóteles Góes-Neto [z], Nicolas G. Bazan [aa], Elrashdy M. Redwan [ab], Murtaza Tambuwala [ac], Vladimir N. Uversky [ad,ae,*]

[a] *Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram, 721140, India*
[b] *Applied Statistics Unit, Indian Statistical Institute, Kolkata, 700108, West Bengal, India*
[c] *Department of Chemistry, College of Art and Sciences, University of South Florida, Tampa, FL, 33620, USA*
[d] *Department of Pharmaceutics and Pharmaceutical Technology, Yarmouk University-Faculty of Pharmacy, Irbid, 566, Jordan*
[e] *Department of Physiology, Michigan State University, East Lansing, MI, 48824, USA*
[f] *PanTherapeutics, Rte de Lavaux 49, CH1095, Lutry, Switzerland*
[g] *Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran*
[h] *Italian Agency for Development Cooperation - Khartoum, Sudan Street 33, Al Amarat, Sudan*
[i] *Department of Food Science, University of Otago, Dunedin, 9054, New Zealand*
[j] *Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, MN, USA*
[k] *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*
[l] *Zoology Department, Faculty of Science, Minia University, El-Minia, 61519, Egypt*
[m] *University of Pittsburgh School of Medicine, Department of Medicine, Division of Hematology/Oncology, UPMC Hillman Cancer Center, Pittsburgh, PA, USA*
[n] *Department of Zoology, Patna University, Patna, 800005, Bihar, India*
[o] *Department of Environmental Health Sciences, Tulane University, New Orleans, LA, 70112, USA*
[p] *Translational Laboratory in Molecular Physiology, Centre for Experimental Surgery, College of Medicine, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil*
[q] *Center for iPS Cell Research and Application, Kyoto University, Japan*
[r] *Biomaterial and Bioengineering Lab, Translational Research Centre San Alberto Magno, Catholic University of Valencia San Vicente M'artir, c/Guillem de Castro 94, 46001, Valencia, Spain*
[s] *School of Engineering and Sciences, Tecnologico de Monterrey, Av. Eugenio Garza Sada 2501 Sur, 64849, Monterrey, Nuevo León, Mexico*
[t] *Department of Molecular Medicine, University of Padova, Via Gabelli 63, 35121, Padova, Italy*
[u] *University of Southern Denmark, Mads Clausen Institute, NanoSYD, Alsion 2, 6400, Sønderborg, Denmark*
[v] *Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, WB, India*
[w] *Departamento de Genética, Ecologia e Evolucao, Instituto de Cîencias Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[x] *Departamento de Ciencias Biologicas (DCB), Programa de Pos-Graduacao em Genetica e Biologia Molecular (PPGGBM), Universidade Estadual de Santa Cruz (UESC), Rodovia Ilheus-Itabuna, km 16, 45662-900, Ilheus, BA, Brazil*
[y] *Laboratório de Bioinformática e Química Computacional, Departamento de Ciências Biológicas, Universidade Estadual do Sudoeste da Bahia (UESB), Jequié, 45206-190, Brazil*
[z] *Laboratório de Biologia Molecular e Computacional de Fungos, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil*
[aa] *Neuroscience Center of Excellence, School of Medicine, LSU Health New Orleans, New Orleans, LA, 70112, USA*
[ab] *King Abdulaz University, Faculty of Science, Department of Biological Science, Saudi Arabia*
[ac] *School of Pharmacy and Pharmaceutical Science, Ulster University, Coleraine, BT52 1SA, Northern Ireland, UK*
[ad] *Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA*

* Corresponding author. Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA.
** Corresponding author.
*** Corresponding author.
E-mail addresses: sarimif@gmail.com (Sk.S. Hassan), lundstromkenneth@gmail.com (K. Lundstrom), vuversky@usf.edu (V.N. Uversky).

https://doi.org/10.1016/j.abb.2022.109124
Received 2 November 2021; Received in revised form 15 January 2022; Accepted 17 January 2022
Available online 24 January 2022

[ae] Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Institutskiy pereulok, 9, Dolgoprudny, 141700, Moscow region, Russia
[af] Network of Immunity in Infection, Malignancy and Autoimmunity (NIIMA), Universal Scientific Education and Research Network (USERN), Stockholm, Sweden

ARTICLE INFO

Keywords:
ORF3a
ORF6
ORF7a
ORF7b
ORF8
ORF10
Pathogenicity
SARS-CoV-2

ABSTRACT

The coronavirus disease 2019 (COVID-19) is caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS- CoV-2) with an estimated fatality rate of less than 1%. The SARS-CoV-2 accessory proteins ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 possess putative functions to manipulate host immune mechanisms. These involve interferons, which appear as a consensus function, immune signaling receptor NLRP3 (NLR family pyrin domain-containing 3) inflammasome, and inflammatory cytokines such as interleukin 1β (IL-1β) and are critical in COVID-19 pathology. Outspread variations of each of the six accessory proteins were observed across six continents of all complete SARS-CoV-2 proteomes based on the data reported before November 2020. A decreasing order of percentage of unique variations in the accessory proteins was determined as ORF3a > ORF8 > ORF7a > ORF6 > ORF10 > ORF7b across all continents. The highest and lowest unique variations of ORF3a were observed in South America and Oceania, respectively. These findings suggest that the wide variations in accessory proteins seem to affect the pathogenicity of SARS-CoV-2.

Executive summary

- SARS-CoV-2 accessory proteins ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 have putative functions to manipulate the host immune system.
- Inflammatory cytokines, such as interleukin 1β (IL-1β), IL-6, and TNF are critical in COVID-19 pathology.
- Extensive heterogeneity was found around six continents for each of the six accessory proteins of all the sequenced SARS-CoV-2 proteomes

1. Introduction

SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2) is the causative agent of the coronavirus disease 2019 (COVID-19) pandemic with an estimated fatality rate of less than 1% [1]. However, Dr Michael Ryan, Executive Director of the Health Emergencies Program at the World Health Organization (WHO), indicated in October 2020 that 760 million people might have been infected by SARS-CoV-2, which gives a hypothetical fatality rate of 0.14%, with approximately one million lives lost. SARS-CoV-2 is a member of the Betacoronavirus (lineage B) genus. The Sarbecovirus subgenus was suggested to had diverged from the lineage of Bat Coronavirus (BatCoV) RaTG13 in 1969 with the 95% highest posterior density interval of the years 1930–2000 [2]. Among previously identified human coronaviruses (HCoVs), Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV) causing the SARS epidemic in 2002–2004 is the closest member to SARS-CoV-2 [2,3]. SARS-CoV possesses eight open reading frames (ORFs), ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8a, ORF8b, and ORF9b, which were suggested to have more intrinsic and secondary roles other than the primary roles described for cellular entry in the viral life cycle [4,5]. For instance, the ORFs are transcribed throughout the second phase of replication by the positive strand subgenomic mRNA using a negative-sense viral RNA template [6]. Thus, due to their intrinsic nature, accessory proteins are not targets for positive-selection such as the extrinsic and primary functional Spike (S) protein containing the receptor-binding domain (RBD) and protease cleavage sites [7]. High-frequency non-synonymous mutations, such as D614G in the S protein detected in clinical SARS-CoV-2 isolates have increased host cell entry via the angiotensin converting enzyme 2 (ACE2) receptor and

```
BCA87362.1      MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFGWLIVGVALLAVFQSAS
MN996532.2      MDLFMRIFTLGTVTLKQGEIKDATPSDSVRATATIPIQASLPFGWLIVGVAFLAVFQSAS
                *********:****************** ***********************:********

BCA87362.1      KIITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLEAPFLYLYALVYFLQSINF
MN996532.2      KIITLKKRWQLALSKGIHFICNLLLLFVTVYSHLLLVAAGLEAPFLYLYALVYFLQSINF
                ****************:**:****************************************

BCA87362.1      VRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPYNSVTSSIVITSGDGTTSPIS
MN996532.2      VRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPYNSVTSSIVITSGDGTTSPIS
                ************************************************************

BCA87362.1      EHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEP
MN996532.2      EHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEP
                ************************************************************

BCA87362.1      EEHVQIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL
MN996532.2      EEHVQIHTIDGSSGVVNPAMEPIYDEPTTTTSVPL
                ******************.****************
```

Fig. 1. ClustalW alignment of SARS-CoV-2 and RaTG13 ORF3 proteins shows 98.5% sequence identity.

Table 1
Total number of six accessory proteins of complete SARS-CoV-2 proteomes.

Proteins	Africa	Asia	Europe	North America	Oceania	South America
ORF3a	280	1175	442	12734	4106	122
ORF6	280	1181	441	12732	4106	122
ORF10	280	1174	442	12733	4106	122
ORF7a	280	1179	440	12723	4106	122
ORF7b	280	1138	436	12568	4106	121
ORF8	280	1172	442	12726	4106	122

Note that all partial accessory proteins and sequences with ambiguous amino acids were excluded from the present study.

Table 2
Ranges and naming of unique sequences (continent-wise) for each accessory protein of SARS-CoV-2.

Continent	ORF3a	ORF6	ORf7a	ORF7b	ORF8	ORF10
Africa	S1 to S7	S1 to S3	S1 to S6	S1 to S2	S1 to S5	S1
Asia	S8 to S85	S4 to S13	S7 to S25	S3 to S9	S6 to S31	S2 to S8
Europe	S86 to S115	S14 to S19	S26	S10 to S11	S32 to S41	S9 to S12
North America	S116 to S442	S20 to S58	S27 to S126	S12 to S30	S42 to S165	S13 to S36
Oceania	S443 to S495	S59 to S69	S127 to S153	S31 to S36	S166 to S186	S37 to S42
South America	S496 to S510	S70 to S72	S154 to S158	S37	S187 to S190	S43 to S44

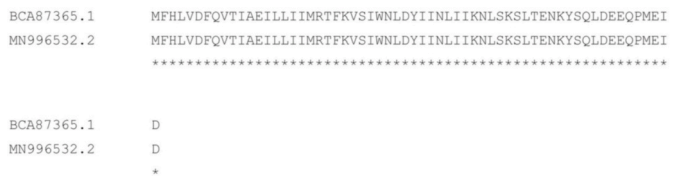

Fig. 2. ClustalW alignment of SARS-CoV-2 (NCBI GenBank ID BCA87365.1) and RaTG13 (NCBI GenBank ID MN996532.2, translated 5′3′ frame 1) ORF6 proteins show 100% sequence identity, despite up to 89 years of genetic diversion.

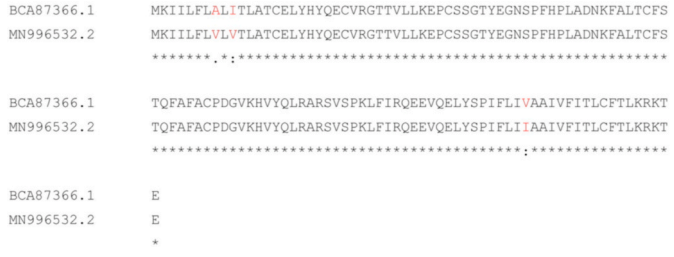

Fig. 3. ClustalW alignment of SARS-CoV-2 (NCBI GenBank ID BCA87366.1) and RaTG13 (NCBI GenBank ID MN996532.2, translated 5′3′ frame 2) The ORF7a proteins show 97.5% sequence identity, despite up to 89 years of genetic diversion.

Fig. 4. ClustalW alignment of SARS-CoV-2 (NCBI GenBank ID BCB15096.1) and Ratg13 (NCBI GenBank ID MN996532.2, translated 5′3′ frame 2) ORF7b proteins shows 97.6% sequence identity, despite up to 89 years of genetic diversion.

Fig. 5. ClustalW alignment of SARS-CoV-2 (NCBI GenBank ID BCA87366.1) and RaTG13 (NCBI GenBank ID MN996532.2, translated 5′3′ frame 2) ORF8 proteins show a 95% sequence identity, despite up to 89 years of genetic diversion.

Fig. 6. ClustalW alignment of SARS-CoV-2 (NCBI GenBank ID BCA87369.1) and RaTG13 (NCBI GenBank ID MN996532.2, translated 5′3′ frame 2) ORF10 proteins show a 97.3% sequence identity, despite up to 89 years of genetic diversion.

transmembrane protease serine 2 (TMPRSS2) [8]. Therefore, due to the intrinsic nature and secondary order in viral transcription, a less selective pressure to induce mutations in accessory proteins is expected. Thus, despite the 19–89 years of estimated genomic divergence between RaTG13 and SARS-CoV-2, the sequence identity between their accessory proteins is very high, being 98.5% for ORF3, 100% for ORF6, 97.5% for ORF7a, 97.6% for ORF7b, 95% for ORF8, and 100% for ORF10. This is indicative of that somehow the direct ancestor of SARS-CoV-2 had been exposed to almost no selection pressure to manipulate its intermediate host immunity for many years until the primary human infection occurred in Wuhan in 2019 (Figs. 1–6) [2]. SARS-CoV-2 and SARS-CoV accessory proteins have differences such as the putative ORF10 protein missing from SARS-CoV and the absence of the ORF3b and ORF9b proteins in SARS-CoV-2 [9,10]. Very little is known about the functions of the accessory proteins of SARS-CoV-2, although crystal or cryo-EM structures were solved for some of them. Examples are given by the Cryo-EM structure of SARS-CoV-2 ORF3a ion channel in lipid nanodiscs (PDB ID: 7KJR) {Kern, 2021 #58}, the X-ray crystal structure of the SARS-CoV-2 ORF7a ectodomain (PDB ID: 7CI3) {Zhou, 2021 #59}, and the crystal structure of the dimeric form of SARS-CoV-2 ORF8 accessory protein (PDB ID: 7JTL) {Flower, 2021 #61}.

The objectives of the present study were to depict the unique variability of all accessory proteins and their possible contributions to virus pathogenicity.

2. Materials and methods

2.1. Data acquisition

Sequences for accessory proteins ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 were downloaded from the complete SARS-CoV-2 proteomes on the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/)(Table 1).

Furthermore, the unique accessory protein sequences were extracted for each continent. The unique protein accessions were renamed for each accessory protein as S1, S2, … etc., as shown in the **Supplementary Tables (S1–S6)**. There were 510, 72, 158, 37, 190, and 44 unique accessory proteins available for ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10, respectively. For each continent, ranges and names of sequences are presented in Table 2.

Statistics of Variants of Accessory Proteins		Africa	Asia	Europe	North America	Oceania	South America
ORF3a	Total	280	1175	442	12734	4106	122
	Unique	7	78	30	327	53	15
	% (continent-wise) among the total	2.50	6.64	6.79	2.57	1.29	12.30
% of unique among all unique sequences		1.67	18.62	7.16	78.04	12.65	3.58
Statistics of Variants of Accessory Proteins		Africa	Asia	Europe	North America	Oceania	South America
ORF6	Total	280	1181	441	12732	4106	122
	Unique (continent-wise)	3	10	6	39	11	3
	% (continent-wise) among the total	1.07	0.85	1.36	0.31	0.27	2.46
% of unique among all unique sequences		5.45	18.18	10.91	70.91	20.00	5.45
Statistics of Variants of Accessory Proteins		Africa	Asia	Europe	North America	Oceania	South America
ORF10	Total	280	1174	442	12733	4106	122
	Unique (continent-wise)	1	7	4	24	6	2
	% (continent-wise) among the total	0.36	0.60	0.90	0.19	0.15	1.64
% of unique among all unique sequences		3.13	21.88	12.50	75.00	18.75	6.25
Statistics of Variants of Accessory Proteins		Africa	Asia	Europe	North America	Oceania	South America
ORF7a	Total	280	1179	440	12723	4106	122
	Unique (continent-wise)	6	19	1	100	27	5
	% (continent-wise) among the total	2.14	1.61	0.23	0.79	0.66	4.10
% of unique among all unique sequences		4.92	15.57	0.82	81.97	22.13	4.10
Statistics of Variants of Accessory Proteins		Africa	Asia	Europe	North America	Oceania	South America
ORF7b	Total	280	1138	436	12568	4106	121
	Unique (continent-wise)	2	7	2	19	6	1
	% (continent-wise) among the total	0.71	0.62	0.46	0.15	0.15	0.83
% of unique among all unique sequences		7.69	26.92	7.69	73.08	23.08	3.85
Statistics of Variants of Accessory Proteins		Africa	Asia	Europe	North America	Oceania	South America
ORF8	Total	280	1172	442	12726	4106	122
	Unique (continent-wise)	5	26	10	124	21	4
	% (continent-wise) among the total	1.79	2.22	2.26	0.97	0.51	3.28
% of unique among all unique sequences		3.40	17.69	6.80	84.35	14.29	2.72

Fig. 7. Number of unique accessory proteins across six continents.

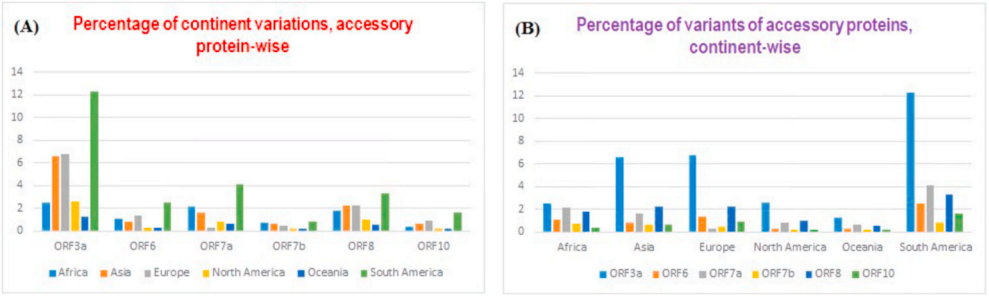

Fig. 8. Bar representations of percentages of continental variations (A), and the percentage of unique accessory proteins (B).

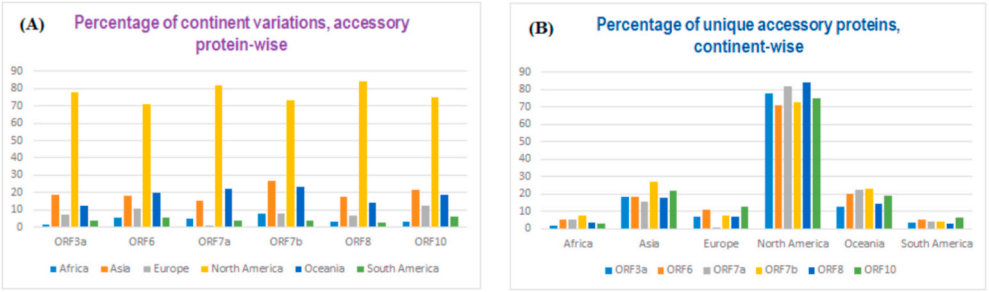

Fig. 9. Quantitative information of the accessory proteins.

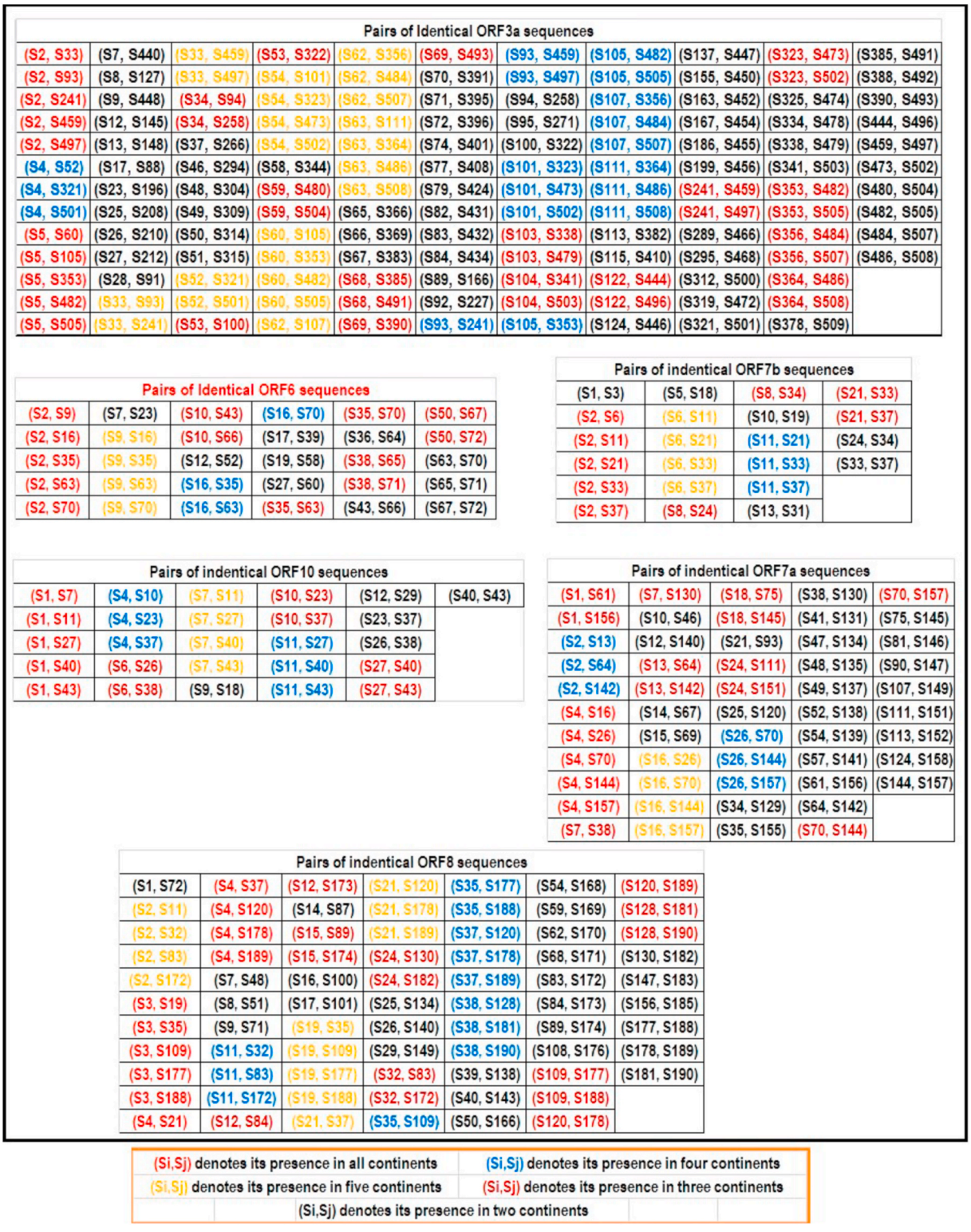

Fig. 10. Identical pairs of accessory protein sequences across all continents.

2.2. Evaluation of the per-residue predisposition of SARS-CoV-2 accessory proteins and their natural variants for intrinsic disorder

Per-residue disorder distribution within the amino acid sequences of SARS-CoV-2 accessory proteins ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 and their natural variants was evaluated by PONDR® VSL2, which is one of the more accurate standalone disorder predictors [11–14]. The per-residue disorder predisposition scores are on a scale from 0.0 to 1.0, where 0.0 indicates fully ordered residues, and 1.0 indicates fully disordered residues. Values above the threshold of 0.5 are considered disordered residues, whereas residues with disorder scores between 0.25 and 0.5 are considered highly flexible, and residues with disorder scores between 0.1 and 0.25 are listed as moderately flexible.

Table 3
List of ORF3a sequences and their distribution over only two continents.

Sequence	Mutation (s)	Present in the continent (s)	Sequence	Mutation (s)	Present in the continent(s)
S7	D2G	Asia and North America	S37	Q57H, A103S	Asia and North America
S8	L15F, Q57H	Asia and North America	S46	L108F	Asia and North America
S9	T32I	Asia and Oceania	S48	W131C	Asia and North America
S12	S40L, Q57H	Asia and North America	S49	L140F	Asia and North America
S13	L41F	Asia and North America	S50	W149L	Asia and North America
S17	V48F	Asia and Europe	S51	T151I	Asia and North America
S23	Q57H, W131C	Asia and North America	S58	DEL (V255), N257D	Asia and North America
S25	Q57H, S166L	Asia and North America	S65	G172V	Asia and North America
S26	Q57H, S171L	Asia and North America	S66	D155Y	Asia and North America
S27	Q57H, T175I	Asia and North America	S67	A99V	Asia and North America
S28	Q57H, S216P	Asia and Europe	S70	K66 N	Asia and North America
Sequence	**Mutation (s)**	**Present in Continent (s)**	**Sequence**	**Mutation (s)**	**Present in Continent (s)**
S71	A54S, Q57H	Asia and North America	S167	V55G	North America and Oceania
S72	A54S	Asia and North America	S186	Q57H, L101F	North America and Oceania
S74	G49V	Asia and North America	S199	Q57H, L140F	North America and Oceania
S77	I35T, Q57H	Asia and North America	S289	G100C	North America and Oceania
S79	D22Y	Asia and North America	S295	V112F	North America and Oceania
S82	G18V, Q57H	Asia and North America	S312	L147F	North America and South America
S83	G18V	Asia and North America	S319	S166L	North America and Oceania
S84	K16 N, Q57H	Asia and North America	S321	S171L	North America and South America
S89	V55F		S325	S177I	North America

Table 3 (*continued*)

Sequence	Mutation (s)	Present in the continent (s)	Sequence	Mutation (s)	Present in the continent(s)
		Europe and North America			and Oceania
S92	Q57H, V237F	Europe and North America	S334	T223I	North America and Oceania
S94	Q57H, D155Y	Europe and North America	S338	T229I	North America and Oceania
S95	Q57H, A99V	Europe and North America	S341	P240L	North America and South America
S100	G172C	Europe and North America	S378	A110S	North America and South America
S113	A39S	Europe and North America	S385	H93Y	North America and Oceania
S115	A33S, Q57H	Europe and North America	S388	H78Y	North America and Oceania
S137	S26L	North America and Oceania	S390	K67 N	North America and Oceania
S155	L46F	North America and Oceania	S444	V13L	Oceania and South America
S163	L53F	North America and Oceania			

2.3. Phylogenetic analysis

In a first step, the SARS-CoV-2 amino acid sequences of each ORF were initially filtered to remove redundant sequences and sequences with low quality (unknown amino acids "X") by using the SeqKit program [15], with the tools fx2tab and rmdup. At this stage, the sequences which presented one or more "X" characters in their composition were removed, as well as redundant sequences (100% identical). Thereafter, amino acid sequences of each ORF group were aligned using the MegaX program [16], applying the MUSCLE algorithm for selection [17]. For all phylogeny estimation the Neighbor-joining method was used, as well as each input alignment was submitted to the phyloXML [18] program, with the multiple alignment inference option, maximum allowed gaps ratio 0.5 and minimum allowed non-gap sequence length 50 with distance calculator Kimura correction. In a last step, phylogenetic trees were analyzed and edited using the phyloXML tool [18].

3. Results and discussion

The essential known features of the six accessory proteins from SARS-CoV-2 are summarized below.

ORF3a protein: The ORF3a is the largest SARS-CoV-2 accessory protein (275 amino acids long). It has 72.4% sequence identity with SARS-CoV ORF3a protein and 98.5% sequence identity with the Bat-CoV RaTG13 ORF3a protein [19,20] (Fig. 1). ORF3a is involved in virulence, infectivity, ion channel activity, morphogenesis, and virus release [21]. In SARS-CoV, ORF3a is a multifunctional protein co-localized with the E,

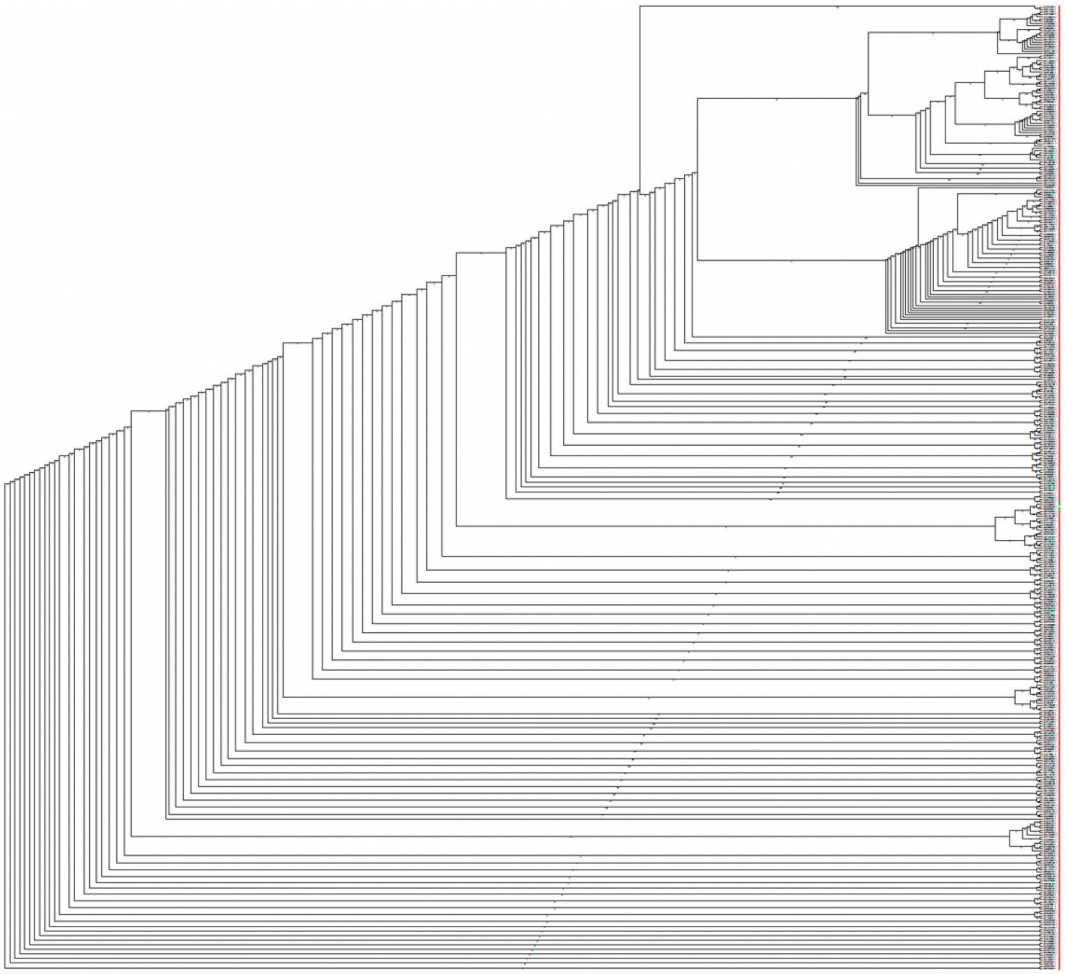

Fig. 11. SARS-CoV-2 ORF3a amino acid phylogeny after group clustering.

M, and S proteins, forming a homo-tetrameric complex as a potassium-ion channel on the host cell membrane during viral assembly [5]. In SARS-CoV-2, the function of the ion-channel proteins (viroporins) ORF3a, ORF8a, and E is critical in tissue inflammation caused by CoVs [6].

Viroporin-mediated lysosomal disruption, and ion-redistribution activate the innate immune signaling receptor NLRP3 (NLR family pyrin domain-containing 3) inflammasome that leads to the expression of inflammatory cytokines such as interleukin 1β (IL-1β), IL-6, and tumor necrosis factor (TNF), causing tissue inflammation during respiratory illness [6] From another pathway, ORF3a interacts with TNF receptor-associated factor (TRAF3) protein with its protein binding domains, which leads to ASC ubiquitination, caspase 1 activation, and IL-1β maturation [22].

Additionally, ORF3a and ORF7a combined with E, S, NSP1 proteins, and MAPK pathway proteins (MAPK8, MAPK14, and MAP3K7) trigger proinflammatory cytokine signaling transcription factors such as STAT1, STAT2, IRF9, and NFKB1 [6]. Additionally, the SARS-CoV-2 ORF3a protein interacts with heme oxygenase-1 (HMOX1) that has a role in heme catabolism and the anti-inflammatory system [6]. ORF3a inhibits cGAS-STING in chicken, mouse and man in a unique fashion and blocks the nuclear accumulation of p65 to inhibit nuclear factor-κB signaling.

Due to more effective innate immune suppression, it may allow more efficient SARS-CoV-2 replication *in vivo*. However, ORF3a was ineffective against the pathways associated with the RIG-I-like receptors (RLRs, which are a family of cytosolic pattern recognition receptors that are essential for detecting viral RNA and initiating the innate immune response) in contrast to the SARS-CoV-2 N protein, which showed strong inhibition of the RLR pathway [23]. The ion channel activity of the SARS-CoV-2 ORF3a, E and M proteins interferes with apoptotic pathways [19]. In a similar scenario, ORF3a of SARS-CoV increases the mRNA expression levels of all three subunits of fibrinogen, thus promoting fibrosis, one of the serious pathogenic aspects of SARS [24]. The expression of NFκB, IL8, and JNK, all involved in the inflammatory responses are also enhanced. Both SARS-CoV-2 ORF3a and ORF3b have showed ability to antagonize type-I interferon activation [25]. Interestingly, potent and durable antibody responses against IFN antagonist SARS-CoV-2 ORF3a, ORF3b, ORF7a and ORF8 proteins have been detected in children [26], which may explain why children are more resistant to SARS-CoV-2 infections [27]. However, it also raises the question, whether the mutations/truncations associated with those accessory proteins will influence the resistance seen in children? Similar to ORF8, ORDF3b is an immune-dominant protein that has been shown to induce high levels of antibody production during SARS-CoV-2 infections [28]. Sequence analysis of ORF3b identified a natural variant with a longer ORF3b reading frame in two patients with severe COVID-19, which enhanced interferon suppression and was potentially linked to viral pathogenesis and severity of COVID-19 [29].

ORF6 protein: SARS-CoV-2 ORF6 is a 61 amino acid long membrane-associated interferon (IFN) antagonist protein. ORF6 interacts with the karyopherin import complex that limits transcription factor STAT1, which down-regulates the IFN pathway [5]. ORF6 is internalized from the plasma membrane into endosomal vesicles. The SARS-CoV-2 ORF6 has a 68.9% sequence identity with the SARS-CoV ORF6 protein and a 100% sequence identity with the BatCoV RaTG13 ORF6 protein [5] (Fig. 2). SARS-CoV ORF6 and ORF3a, in association

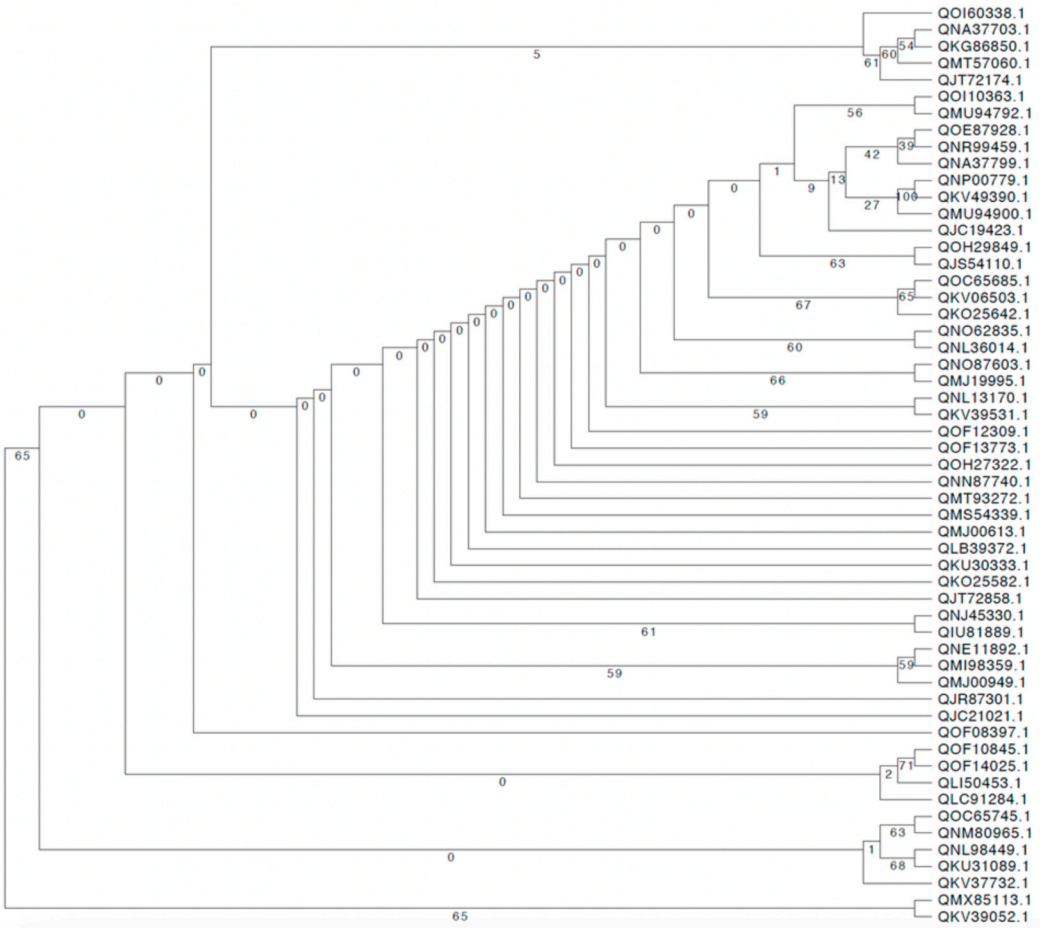

Fig. 12. SARS-CoV-2 ORF6 amino acid phylogeny after group clustering. Phylogenetic analysis identified four well-defined groups.

with other proteins such as M, NSP1 and NSP3 inhibit IRF3 signaling, repress interferon expression and stimulate the degradation of IFNAR1 and STAT1 [6].

The SARS-CoV-2 ORF6 interacts with the NSP8 protein, and it can increase early infection at a low multiplicity with an increase in RNA polymerase activity [30]. It has been reported that ORF6 and ORF8 can inhibit the type-I IFN signaling pathway [30]. The ORF6 protein with the lysosomal targeting motif (YSEL) and diacidic motif (DDEE) induces intracellular membrane rearrangements resulting in a vesicular population and endosomal internalization of viral protein into infected cells increasing replication [31].

ORF7a and ORF7b proteins: ORF7a, a 121 aa type I transmembrane protein, interacts with SARS-CoV-2 structural proteins M, E, and S, which are essential for viral assembly. Hence, ORF7a is involved in viral replication, and virion-associated ORF7a protein may function during early infection. It has an 85.2% sequence identity with the SARS-CoV ORF7a protein and has a 97.5% sequence identity with BatCoV RaTG13 ORF7a protein [5] (Fig. 3).

ORF7a interacts with the SARS-CoV-2 M, E and S structural proteins, which are essential for viral assembly, and hence ORF7a is involved in viral replication, and virion-associated ORF7a protein may function during early infection [32–34]. ORF7a induces pro-inflammatory cytokines and chemokines, such as IL-8 and RANTES [5]. SARS-CoV-2 ORF7a in combination with the E protein activates apoptosis by suppressing anti-apoptotic proteins [6]. While ORF7b is a 43 aa protein found in association with intracellular viral particles, it is also present in purified virions in the Golgi compartment. The SARS-CoV-2 ORF7b has an 85.4% sequence identity with SARS-CoV ORF7b protein and has a 97.6% sequence identity, with BatCoV RaTG13 ORF7a protein [5] (Fig. 4).

ORF7b is found associated with intracellular viral particles and purified virions. To date, there is extraordinarily little experimental evidence to support a role for ORF7a or ORF7b in SARS-CoV-2 replication [32].

ORF8 protein: ORF8 is a unique 121 aa long accessory protein in SARS-CoV-2, and it stands out by being poorly conserved among other CoVs, accordingly showing structural changes suggested to be related to the ability of virus spread [35]. ORF8 sequences of SARS-CoV-2 and RaTG13 share a 95% amino acid identity (Fig. 5).

SARS-CoV-2 ORF8 interacts with the major histocompatibility complex (MHC) class-I molecules and down-regulates their surface expression on various cell types [36]. It has been reported earlier that inhibition of ORF8 could be a strategy to improve the special immune surveillance and to accelerate the eradication of SARS-CoV-2 *in vivo* [37].

ORF10 protein: The 38 aa long ORF10 accessory protein has been reported to be unique for SARS-CoV-2 containing eleven cytotoxic T lymphocyte (CTL) epitopes of nine amino acids each in length, across various human leukocyte antigen (HLA) subtypes [38,39]. ORF10 negatively affects the antiviral protein degradation process through its interaction with the Cul2 ubiquitin ligase complex [6]. The ORF10 protein is missing in SARS-CoV, but SARS-COV-2 ORF10 and RaTG13 ORF10 have a 97.3% sequence identity [40] (Fig. 6).

For every continent, the total number of accessory proteins and the total number of unique sequences with respective percentages are presented in Fig. 7. In summary, for all six continents, the total number of unique ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 accessory protein sequences are 419, 55, 122, 26, 147, and 32, respectively

Table 4
List of ORF7a sequences and their distribution over only two continents.

Sequence	Mutation (s)	Present in the continent(s)	Sequence	Mutation (s)	Present in the continent(s)
S10	V71I	Asia and North America	S49	S81L	North America and Oceania
S12	Q94H	Asia and Oceania	S52	S83L	North America and Oceania
S14	L116F	Asia and North America	S54	V93F	North America and Oceania
S15	T120I	Asia and North America	S57	L96F	North America and Oceania
S21	C67Y	Asia and North America	S61	P99L	North America and South America
S25	A13T	Asia and North America	S81	E95Q	North America and Oceania
S34	T28I	North America and Oceania	S90	H73Y	North America and Oceania
S35	V29L	North America and South America	S107	H47Y	North America and Oceania
S41	T39I	North America and Oceania	S113	P34S	North America and Oceania
S47	Q76H	North America and Oceania	S124	A8V	North America and South America
S48	R79C	North America and Oceania			

Table 5
List of ORF8 sequences and their distribution over only two continents.

Sequence	Mutation (s)	Present in the continent(s)	Sequence	Mutation(s)	Present in the continent(s)
S1	V33F	Africa and North America	S40	P38S	Europe and North America
S7	T11I	Asia and North America	S50	T11K	North America and Oceania
S8	T12 N	Asia and North America	S54	S21 N	North America and Oceania
S9	V32L	Asia and North America	S59	S24L, DEL (DS)66–67, K68E	North America and Oceania
S14	G66C	Asia and North America	S62	S24L	North America and Oceania
S16	P93L	Asia and North America	S68	Q27K	North America and Oceania
S17	L95F	Asia and North America	S108	V114	North America and Oceania
S25	D63 N	Asia and North America	S130	A65V	North America and Oceania
S26	A51V	Asia and North America	S147	P36S	North America and Oceania
S29	D34G	Asia and North America	S156	G8R	North America and Oceania
S39	A55V	Europe and North America			

(Supplementary Figure S1). Furthermore, the percentage of unique sequences on each continent among all available accessory proteins are also enumerated (Fig. 7).

The percentages of each accessory protein across the six continents are presented as bar diagrams in Fig. 8. The following observations were drawn from Fig. 8. Across all continents, the decreasing order of percentage of unique variations in the accessory proteins was observed as follows: ORF3a > ORF8 > ORF7a > ORF6 > ORF10 > ORF7b. The highest and lowest unique variations of ORF3a were observed in South America and Oceania, respectively. In addition, the highest percentage (statistically significant) of unique variations in each accessory protein was observed in South America. The lowest percentage of unique variations among ORF3a, ORF6, ORF7b, and ORF8 was observed in Oceania. It is worth noticing that the smallest number of unique variations of ORF7b and ORF7a was seen in North America and Europe, respectively. It is further noted that in Europe, the lowest variations among all accessory proteins were found in ORF7a. The smallest percentage of unique ORF10 variations was found in Oceania. With regards to the total unique variations across all accessory proteins of SARS-CoV-2, the decreasing order would be in South America > Asia > Europe > Africa > North America > Oceania.

ORF3a possessed the highest significant amount of unique variations across all six continents, while ORF10 showed the lowest variations in Africa, Asia, and Oceania. The lowest unique variations of ORF7b were observed in North America and South America.

The percentage of unique accessory proteins among all unique sequences obtained across the six continents is represented as bar diagrams in Fig. 9.

Among all available unique variations of the six accessory proteins of SARS-CoV-2, North America and South America exhibited the highest and lowest percentage of each accessory protein variation, respectively. The smallest number of unique variations of ORF3a, ORF6, and ORF10 were noticed in Africa. On the other hand, South America showed the lowest number of unique ORF6, ORF7a, ORF7b, and ORF8. Regarding ORF7b, the highest number of unique variations compared to the rest of the accessory proteins were observed in Africa, Asia, and Oceania. Furthermore, the highest percentage (84.35%) and lowest (0.82%) of unique variations of ORF8 and ORF7a (among all accessory proteins) were found in North America and Europe, respectively.

Fig. 10 represents the continent-wise lists of identical sequences for each accessory protein. The following observations were made for each accessory protein based on the data shown (Fig. 10).

ORF3a: Note that the mutations described below were determined based on the Wuhan ORF3a sequence (YP 009724391). There were only two ORF3a sequences (marked in red), S2 (Africa, QOI60359) and S5 (Africa, QOI60335), which were present on all six continents.

Note that the S2 (Africa-ORF3a) was identical with ORF3a (YP 009724391) from Wuhan, China. The other sequence, S5, is different from ORF3a (YP 009724391) by one missense mutation Q57H, a strain-determining mutation [41]. It was found that the ORF3a sequence S54 (Asia: QKK14624) possesses the single T175I mutation and is present on all continents except in Africa. The ORF3a sequences S62 (Asia: QMJ01306) and S63 (Asia: QJQ04482) possessed a single mutation each, G251V and G196V, respectively, compared to the Wuhan ORF3a (YP 009724391). These two sequences were present in Asia, Europe, North America, Oceania, and South America. The ORF3a sequence S4 (Africa: QLQ87565) has the single S171L mutation found on four continents, excluding Europe and Oceania. Two mutations, Q57H and D155Y, in sequence S34 (Asia), were present only on three continents, Asia, Europe, and North America. Sequence S53 (Asia) with the G172C mutation was found in Asia, Europe, and North America.

The deletion mutation V255 occurred in S59 (Asia), which was found

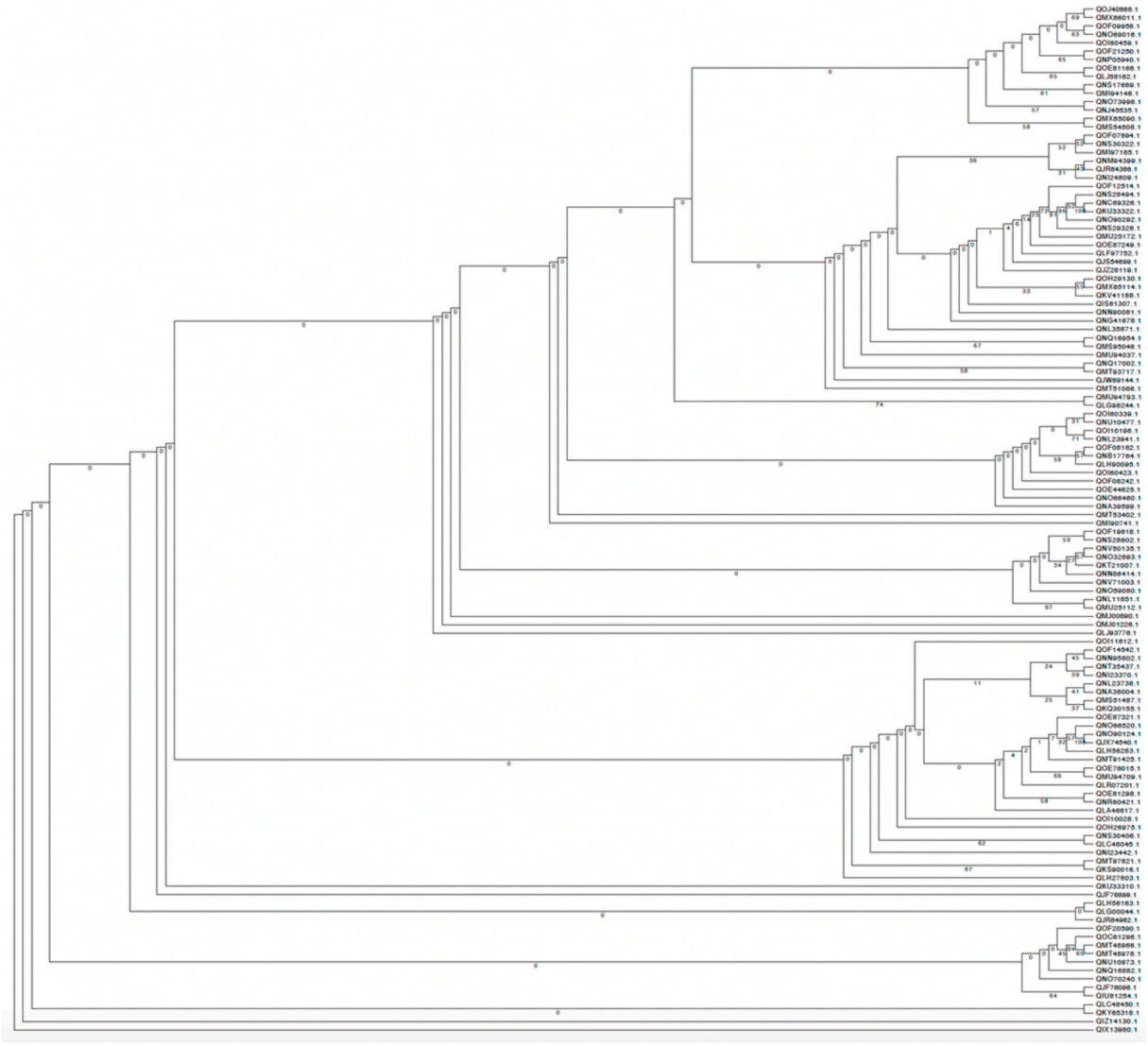

Fig. 13. SARS-CoV-2 ORF7a amino acid phylogeny after group clustering. Two well-defined groups can be identified.

in Asia, Oceania, and South America. S68 (Asia) and S69 (Asia) possessed two mutations, H93Y and K67 N, respectively. These two ORF3a variants have been detected only on three continents, Asia, North America, and Oceania.

The ORF3a sequence S103 containing the single T229I mutation is present only on three continents, Europe, North America, and Oceania. Another sequence, S104, with the P240L mutation has been detected only in Europe, North America, and South America. The V13L mutation was found in sequence S122 (ORF3a, North America) and is present on three continents, Oceania, North America, and South America. Further, there were 57 unique ORF3a variants detected only on two continents as listed in Table 3.

Fig. 11 represents a phylogenetic tree for SARS-CoV-2 ORF3a proteins. This ORF3a tree was composed by the alignment of 419 sequences, and its resultant phylogeny shows that there are no well-defined patterns for the grouping of sequences, as well as it is possibly not showing evolutionary relationships, but random mutation events. These results show that ORF3 does not seem to represent a target for selection pressure and, therefore, phylogenetic analysis of this protein does not provide noticeable grounds for making associations or evolutionary and/or lineage relationships between the strains.

ORF6: Note that the mutations described below were determined based on the Wuhan ORF6 sequence (YP 009724394). The sequence S2 (ORF6, Africa) was identical with YP 009724394 (China, Wuhan) ORF6, and this sequence was present on all six continents, whereas the ORF6 sequence, S10 (ORF6, Asia) with only the D53Y mutation, was found only in Asia, North America, and Oceania. The ORF6 sequences S38 (ORF6, North America) and S50 (ORF6, North America) possess a single mutation each, D2L and I33T, respectively, found on three continents, North America, Oceania, and South America. The ORF6 unique variant S7 (ORF6, Asia) possesses the E13D mutation found only in Asia and North America. The ORF6 sequence S12 (ORF6, Asia) possess a set of deletions," FKVSIWNLD" (22–30 aa), and it appeared in Asia and North America only. The sequence S17 (ORF6, Europe) had the D61Y mutation, and it was found in Europe and North America. In addition, a single mutation H3Y occurred in S19 (ORF6, Europe), which was present in Europe and North America. The ORF6 sequence S27 (ORF6, North America) containing the W27L mutation was found in North America and Oceania only. Furthermore, the sequence S36 (ORF6, North America) with the D61H mutation was present in North America and Oceania only.

Fig. 12 represents a phylogenetic tree for the ORF6 protein. This tree was constructed by the alignment of 55 sequences, and it was possible to identify four very distinct groups. On the other hand, most sequences did not present a clear grouping.

ORF7a: Mutations are based on the Wuhan ORF7a sequence (YP

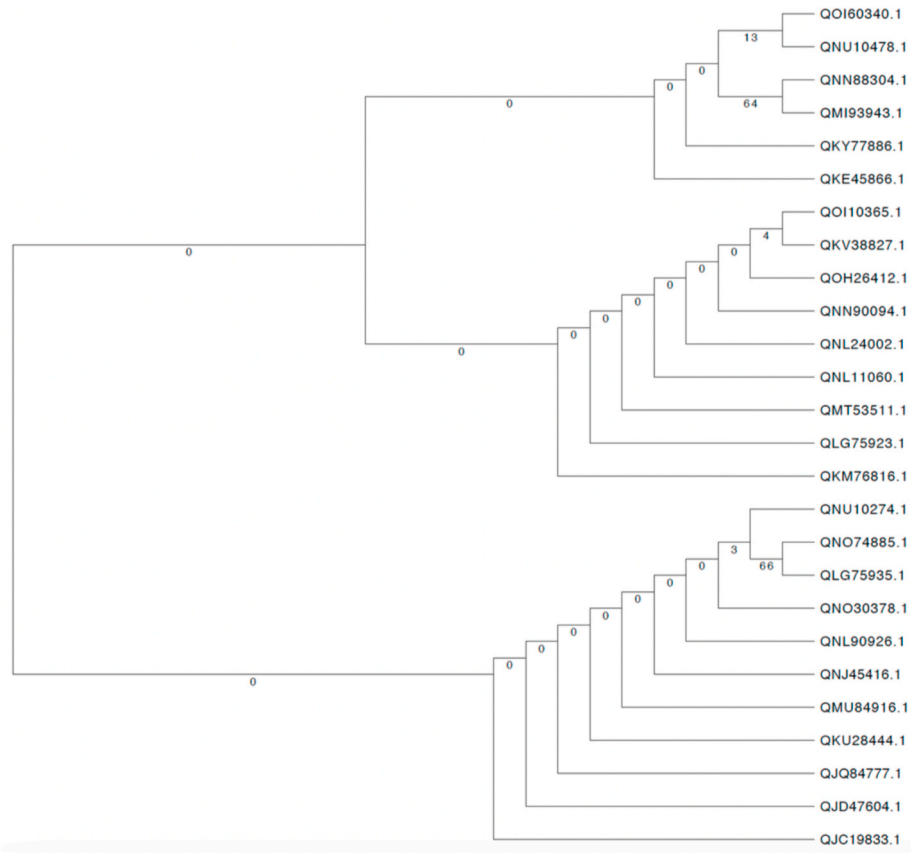

Fig. 14. SARS-CoV-2 ORF7b amino acid phylogeny after group clustering. Analysis identified three well-defined groups.

009724395). The Wuhan ORF7a sequence YP 009724395 was found on all continents. V104F was found in S2 (ORF7a, Africa) in Africa, Asia, North America, and Oceania. The sequence S1 (ORF7a, Africa) had the P39L mutation, which was found in Africa, North America, and South America. S37F was found in the sequence S7 (ORF7a, Asia) in Asia, North America, and Oceania. The sequence S18 (ORF7a, Asia) has the A105V mutation found across Asia, North America, and Oceania. G38V was found in S24 (ORF7a, Asia) in Asia, North America, and Oceania. Also, there were 21 unique ORF7a variants present only on two continents. All mutations are listed in Table 4.

The phylogenetic analysis for the 122 amino acid sequences of the ORF7a revealed the presence of two clear groups, with the first group containing most of the sequences. On the other hand, four non-grouped sequences were found as well (Fig. 13).

ORF7b: Here, all mutations are accounted based on the Wuhan ORF7b sequence (YP 009725318). The sequence S2 (ORF7b, Africa) (identical to Wuhan ORF7b (YP 009725318)) was found on all six continents. It was found that only the C41F mutation was present in S8 (ORF7b, Asia), which appeared in Asia, North America, and Oceania. The sequence S1 (ORF7b, Africa) had the single mutation S5L, present in Africa and Asia. The sequence S5 (ORF7b, Asia) had the mutation S31L, and this sequence was found on two continents, Asia and North America only. L32F occurred in the sequence S10 (ORF7b, Europe), present in Europe and North America. Furthermore, the sequence S13 had the mutation L4F, and this sequence was found in North America and Oceania.

For the ORF 7b proteins, phylogenetic analysis was performed using 26 amino acid sequences. Fig. 14 shows that the corresponding phylogenetic tree has three well-defined groups. In this phylogeny, an evolutionary proximity relationship between the sequences can be verified (Fig. 14).

ORF8: Mutations described below are determined regarding the Wuhan ORF8 sequence (YP 009724396). It was observed that the Wuhan ORF8 YP 009724396 sequence was found on every continent. Also, another sequence present on every continent was the single mutation L84S. The single mutaion V62L was observed in the sequence S2 (ORF8, Africa), which was found on all continents except South America, whereas the ORF8 sequence S38 (Europe) possessed the single mutation A65S, and the sequence was found in North America, Oceania, and South America. Further, the V62L and L84S mutations were observed in S12 (ORF8, Asia) in Asia, North America, and Oceania. The sequence S15 (ORF8, Asia) contained the mutation S67F, which was found in Asia, North America, and Oceania. The ORF8 sequence S24 (Asia) possessed the single mutation A65V, which was found in Asia, North America, and Oceania.

In the ORF8 phylogenetic analysis, we used 147 amino acid sequences. Fig. 15 shows the presence of three well-defined groups. On the other hand, many sequences were not grouped, and did not present well-defined branches.

ORF10: Mutations are based on the Wuhan ORF10 sequence (YP 009725255). The Wuhan ORF10 (YP 009725255) was identical with S1 (ORF10, Africa), and it was found on every continent. The ORF10 sequence S6 (ORF10, Asia) had the mutation L37F, and the sequence was present in North America and Oceania only. The V30L mutation was only found in the ORF10 sequence S10 (Europe), which appeared in Europe, North America, and Oceania. The sequence S9 (ORF10, Europe) had the mutation S23F, and it was found in Europe and North America. The mutation D31Y appeared in the S12 sequence (ORF10, Europe), which was found in Europe and North America only.

The results for the ORF10 phylogenetic analysis included 32 sequences and showed four groups, the first with eight sequences, the second with 16, and the last two groups with four sequences each (Fig. 16).

Concluding this section, one need to keep in mind that the phylogeny

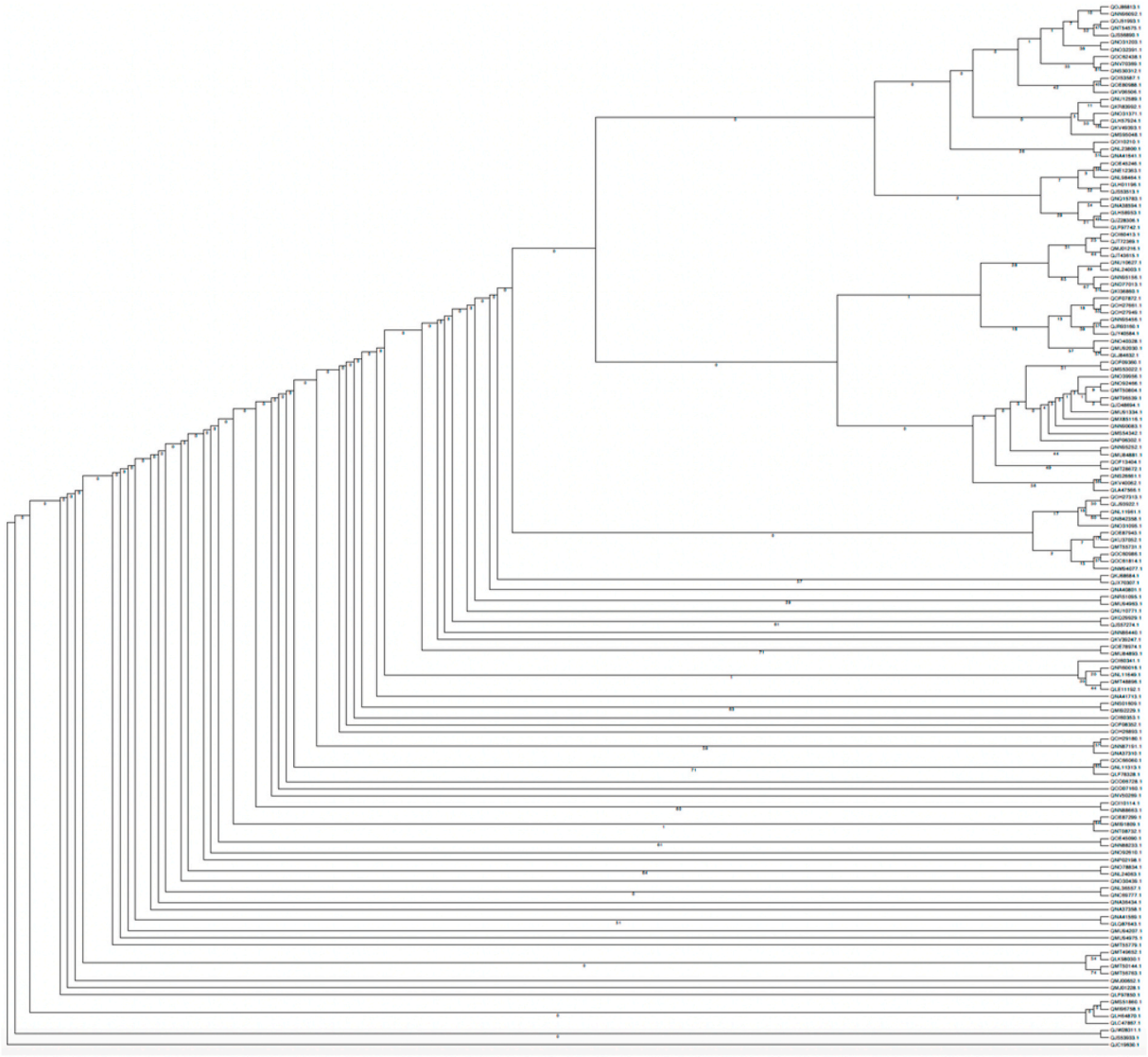

Fig. 15. Phylogenetic analysis of SARS-CoV-2 ORF8 protein identified three well-defined groups.

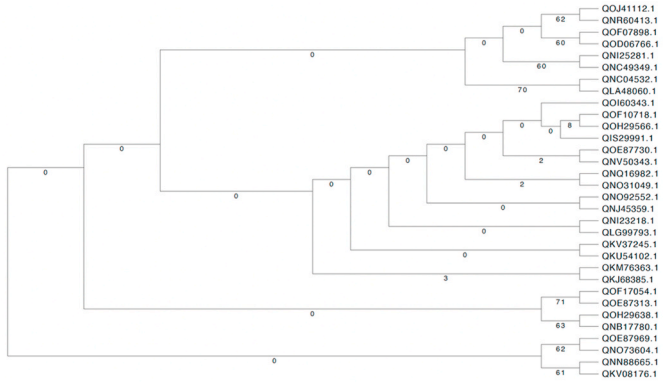

Fig. 16. SARS-CoV-2 ORF10 amino acid phylogenetic analysis identified four well-defined groups.

results are only suggestive and can be used for finding new possibilities to search for other genes in association with the vaccine and/or drug development, which typically works best with well-defined strain clades (see Table 5).

3.1. Featuring uniqueness of the accessory proteins

Here, certain basic descriptive statistics (mean, variance, lower bound, upper bound, and range) were employed to describe the variability of the percentage of the predicted intrinsically disordered residues (PPIDRs), molecular weight (MW), and isoelectric point (pI) of all the unique variants of all accessory proteins (Table 6). The zigzag behavior of the plots of PPIDRs, MW, and pI depicts wide variability of variants for each accessory protein (Supplementary Figures S2–S41).

The following observations were made based on the data shown in Table 6. The amount of total dispersion (based on range) of the percentage of PPIDR and MW of ORF6 variants was highest, whereas the highest amount of total dispersion of pI was observed for ORF10. The smallest amounts of total dispersions of the percentage of PPIDR, MW, and pI were found for ORF3a, ORF10, and ORF7b, respectively. The broad range and variance of the MW values of the unique ORF3a, ORF7a, ORF8, and ORF10 variants imply the wide variability of each set of ORF3a, ORF7a, ORF8, and ORF10 although range and variance of PPIDR and pI were not widely spread. In the case of the unique variance of ORF6, the range and variance of MW and percentage of PPIDR were found to be large, which implied the wide quantitative differences among the unique ORF6 variants. Furthermore, a moderately broad

Table 6
Descriptive statistics of PPIDR, MW, and IP of unique accessory proteins of SARS-CoV-2.

PPIDR of unique accessory proteins of SARS-CoV-2based on PONDR® VSL2					
Accessory proteins	Mean	Variance	Lower bound	Upper bound	Range
ORF3a	4.756	0.2328	2.91	7.64	4.73
ORF6	25.74	74.69	21.31	87.5	66.19
ORF7a	3.51	0.5716	2.48	7.29	4.81
ORF7b	44.663	10.527	37.21	51.16	13.95
ORF8	9.125	1.285	5.6	13.45	7.85
ORF10	18.67	5.0691	13.16	23.68	10.52
MW of unique accessory proteins of SARS-CoV-2					
Accessory proteins	***Mean***	***Variance***	***Lower bound***	***Upper bound***	***Range***
ORF3a	31123	17917.58	29187	31270	2083
ORF6	7171.03	371714.6	2881.205	7542.84	4661.635
ORF7a	13673.4	150719.4	10874.515	14328.65	3454.135
ORF7b	5173.02	2651.26	5033.005	5224.22	191.215
ORF8	13841.4	21411.43	12608.465	14431.55	1823.085
ORF10	4446.53	1173.801	4389.085	4509.285	120.2
pI of unique accessory proteins of SARS-CoV-2					
Accessory proteins	***Mean***	***Variance***	***Lower bound***	***Upper bound***	***Range***
ORF3a	5.9127	0.0278	5.2349	6.5881	1.3532
ORF6	4.4013	0.057	3.8436	5.7589	1.9153
ORF7a	8.0932	0.0434	6.7486	8.5946	1.846
ORF7b	3.9519	0.0063	3.6379	4.1442	0.5063
ORF8	5.6368	0.1223	4.7442	6.8829	2.1387
ORF10	8.2415	0.6857	6.0601	9.2043	3.1442

range and variance associated with the percentage of PPIDR and MW of ORF7a variants imply their moderate variability.

In line with the previously reported data, Figs. 17 and 18 and Table 6 show that all SARS-CoV-2 accessory proteins contain different levels of intrinsic disorder. In fact, based on their overall disorder predispositions, these proteins can be arranged as follows: ORF8 < ORF3a < ORF7a < ORF10 < ORD6 < ORF7b, where the difference in the overall intrinsic disorder predisposition between these proteins can be as high as 6-7-fold (compare data for ORF8 and ORF7b in Fig. 17). Furthermore, disorder predispositions of these proteins are sensitive to the mutations found in their natural variants. For example, Fig. 17 represents the effect of mutations in the natural variants on the overall disorder predisposition of accessory proteins and shows that the whole-protein disorder-related parameters, PPIDR and mean disorder score (MDS), can be dramatically changed by mutations. The largest variability of mutation-induced change in intrinsic disorder propensity is observed for ORF10 and ORF6 (see Fig. 17).

Next, we looked at the effect of natural variants on local intrinsic disorder predisposition. Results of this analysis are shown in Fig. 18, which represents the per-residue disorder profiles generated by PONDR® VSL2 for all the proteins analyzed in this study. Fig. 18 generally supports the observation that intrinsic disorder predispositions could vary significantly between the natural variants of each individual accessory protein. Importantly, the largest mutation-induced variability is observed within the disordered or flexible regions of these proteins (i.e., regions characterized by the predicted disorder scores exceeding the 0.5 threshold and regions with disorder scores between 0.15 and 0.5). This is an important observation suggesting that the natural variability of SARS-CoV-2 accessory proteins is shaping their structural flexibility.

SARS-CoV-2 is the first HCoVs with pandemic capacity due to its highly contagious nature deriving from the structural differences in its S protein, such as a flat sialic acid-binding domain, tight binding to its entry ACE2 receptor, and capacity to be cleaved by furin protease [50]. Based on more than 355 million confirmed cases of COVID-19 and additionally a large number of asymptomatic cases, SARS-CoV-2 is a

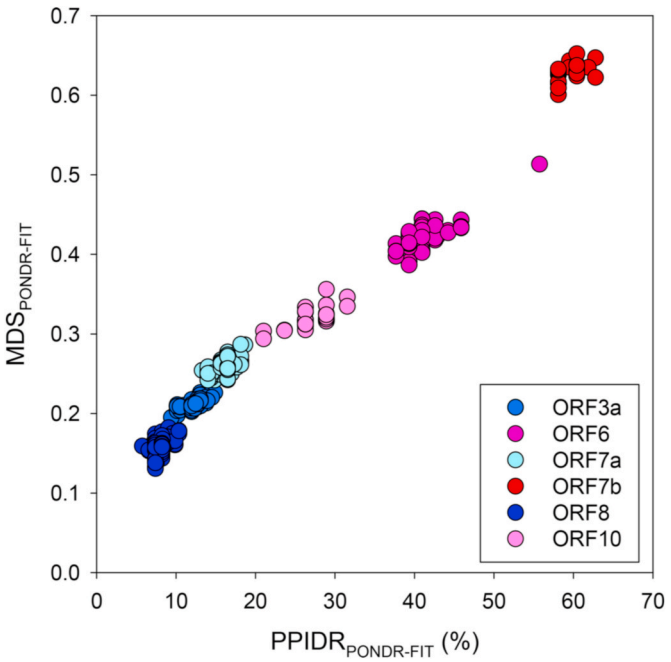

Fig. 17. Effect of mutations observed in unique natural variants of the SARS-CoV-2 accessory proteins on their overall intrinsic disorder predisposition evaluated in terms of percent of predicted intrinsically disordered residues (PPIDR) and mean disorder score (MDS). These data were generated using the PONDR® FIT [42] algorithm, which is a meta predictor that combines outputs of six predictors of intrinsic disorder, PONDR® VLXT [43], PONDR® VSL2 [44,45], PONDR® VL3 [46], FoldIndex [47], IUPred [48], and TopIDP [49]. PONDR® FIT is moderately more accurate than each of its component predictors [42]. For each mutant, the predicted percentage of intrinsically disordered residues (PPIDR) and mean disorder score (MDS) were calculated based on the outputs of this per-residue disorder predictors. Here, PPIDR in a query protein represents a percentage of residues with disorder scores exceeding 0.5. In this study, protein residues and regions were classified as disordered or flexible if their predicted disorder scores were above 0.5, or ranged between 0.15 and 0.5, respectively.

highly contagious, but relatively weak pathogen considering the ratio of the number of patients with severe infections associated with the multiple organ dysfunction to the total number of infected [6], or relatively low mortality rate (~2.2%). The host immunity modulated by the SARS-CoV-2 accessory proteins could be responsible at least for some of these pathological features.

Based on various mutations of accessory proteins, SARS-CoV-2 has had very little selective pressure to tackle host immunity in nature after diverging with BatCoV RaTG13 19–89 years ago [2]. The genomic stability of the relatively large RNA genomes (around 30,000 nucleotides) of SARS-CoV-2, as other CoVs, is protected by proofreading proteins, such as 3′-5′ exonuclease non-structural protein 14 (NSP14) that assists RNA synthesis with a unique RNA proofreading function [51]. Muller's ratchet effect explains the extinctive effect of high mutation rates of asexual organisms such as viruses potentially leading to the accumulation of deleterious mutations in an irreversible manner [52]. Therefore, SARS-CoV-2 repairs its mutations to preserve its genomic stability as mutations can lead to pathological fitness losses or viral extinction [52]. However, there is a balance governed by genomic repair mechanisms such as NSP14, and viruses that require a certain degree of mutations to gain novel traits such as emergence transmission into zoonotic hosts [52]. For instance, a 29-nucleotide deletion mutation in the SARS-CoV ORF8 gene, was associated with a less pathogenic strain [52]. Similarly, SARS-CoV-2 variants with a 382-nucleotide deletion in ORF8, showed only mild symptoms in COVID-19 patients, as they did not require supplemental oxygen [52].

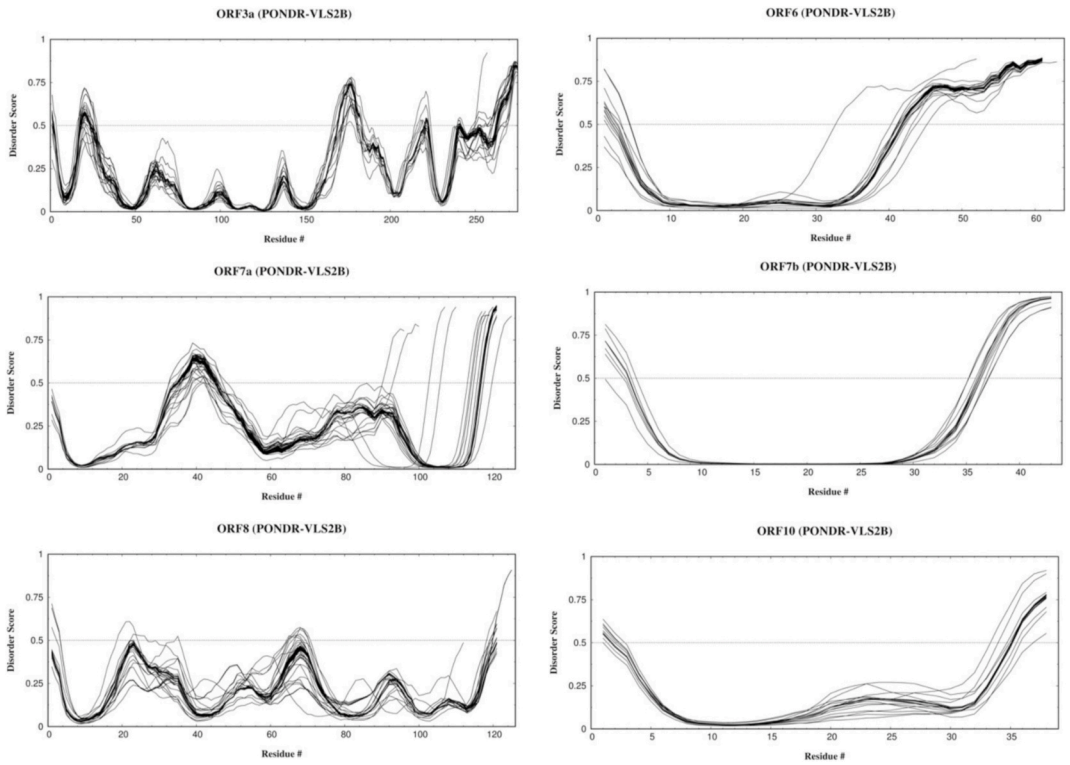

Fig. 18. Per-residue intrinsic disorder profiles generated for the SARS-CoV-2 accessory proteins and their natural variants by PONDR® VSL2, which systematically shows good performance in various comparative analyses, including recently conducted Critical assessment of protein intrinsic disorder prediction (CAID) experiment, where this tool was recognized as #3 predictor of 43 evaluated methods [31].

Only one variant identical to the Wuhan sequence (NC 045512) of each of the accessory ORF6, ORF7a, ORF7b, and ORF10 proteins was present on all continents. Most of the ORF3A variants with the prevalent non-synonymous amino acid substitutions (V13L, T14I, L46F, A54S, Q57H, S58 N, K75 N, A99V, L108F, R126S, G172V, G196V, F207L, T223I, G251I, G252V, N257S, and Y264C) possess a single point mutation [53,54]. Ten of these mutation sites occur within the transmembrane (TM) domain of ORF3a. Four of these variants contain the mutation Q57H paired with another amino acidic change (A99V, S58 N, Y264C, or G172V). Only two variants of ORF3a, differed by the clade/strain determining single mutation Q57H, were found on all six continents [41], and V13L, Q57H+A99V, G196V, and G252V were the most frequent mutations [54]. When Q57H and G251V (ORF3a) are combined with S19L and R203K/G204R in the nucleocapsid, these four mutations cause a dramatic change in viral protein structures [55]. In addition to being predominating in North America [53,56], some ORF3a variants were found on all six continents. This can be associated with virus evasion of the immune system leading to induction of cytokine, chemokine, and interferon-stimulated gene expression in primary human respiratory cells [25,57]. These dominating mutational effects are not limited to the modulation of the efficiency of viral pathogenesis, disease severity, and patient outcomes due to aggravation of the host immunity [21,53,58]. It may also play a role in viral ion-channel formation, viral particle loads, and virus release [21]. The precise roles of natural and/or variants of various SARS-CoV-2 ORFs on the outcome of COVID-19 patients are rather controversial [59] and need a more in depth analysis. Also, in ORF8, only two unique variants, differed by a strain determining single mutation L84S, appeared on all continents. So, the maximally intersecting family of variations across all accessory proteins turned out to be the smallest. These findings confirmed that all other variants of accessory proteins were due to demographic and environmental constraints.

It was found that most of the unique variants of accessory proteins differed from the corresponding Wuhan accessory proteins by a single mutation, although basic descriptive statistics unfolded their respective wide variability. New variants of each accessory protein have been found in recent days and will continue to be discovered in the future. Significant amounts of unique variants of each accessory protein with wide variability might significantly contribute to the pathogenicity of SARS-CoV-2.

Therefore, our firm conviction that naturally weakened stability (if achievable) of SARS-CoV-2 seems to be a far reachable goal, which needs to address the dangers of the present pandemic scenario. Also, unique accessory protein variants across individual continents would all be expected to be mixed, while international travels will resume without strict protective measures and restrictions. In this regard, it is our (SACRED, Self-Assembled COVID-19 Research & Education Directive, consisting of international experts in mathematics, physics, computer science, bioinformatics, nanotechnology, structural biology, molecular biology, immunology, and virology) strong recommendation to governmental and non-governmental organizations to take necessary measures to mitigate the spread of COVID-19.

4. Future perspective

In comparison to either SARS or MERS alone or combined, COVID-19 has caused more illness and death. CoVs can similarly trigger spreads and outbreaks in the coming years with different waves of variants as part of increased globalization. Broad spectrum genomics experiments should be used for the identification of possible genetic factors involved in COVID-19 development. Although costly and complicated, more genomics studies are required to assess the effect of host genomics and genetics on immune responses to CoV. Furthermore, understanding the progression and geographical location of SARS-CoV-2 viral genomics and genetics in the context of frequency and quantity of emerging viral variants and their association with viral infectivity, transmissibility, and clinical manifestation are issues to be addressed in future research and development programs.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.abb.2022.109124.

References

[1] E. Petersen, M. Koopmans, U. Go, D.H. Hamer, N. Petrosillo, F. Castelli, M. Storgaard, S. Al Khalili, L. Simonsen, Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics, Lancet Infect. Dis. 20 (9) (2020) e238–e244.

[2] M.F. Boni, P. Lemey, X. Jiang, T.T. Lam, B.W. Perry, T.A. Castoe, A. Rambaut, D. L. Robertson, Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic, Nat. Microbiol. 5 (11) (2020) 1408–1417.

[3] K. Ohnuki, M. Yoshimoto, H. Fujii, Radiological protection and biological COVID-19 protection in the nuclear medicine department, Eur. J. Nucl. Med. Mol. Imag. 48 (1) (2021) 6–8.

[4] M.L. Dediego, L. Pewe, E. Alvarez, M.T. Rejas, S. Perlman, L. Enjuanes, Pathogenicity of severe acute respiratory coronavirus deletion mutants in hACE-2 transgenic mice, Virology 376 (2) (2008) 379–389.

[5] R. Giri, T. Bhardwaj, M. Shegane, B.R. Gehi, P. Kumar, K. Gadhave, C.J. Oldfield, V. N. Uversky, Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses, Cell. Mol. Life Sci. 78 (4) (2021) 1655–1688.

[6] M. Ostaszewski, A. Mazein, M.E. Gillespie, I. Kuperstein, A. Niarakis, H. Hermjakob, A.R. Pico, E.L. Willighagen, C.T. Evelo, J. Hasenauer, F. Schreiber, A. Drager, E. Demir, O. Wolkenhauer, L.I. Furlong, E. Barillot, J. Dopazo, A. Orta-Resendiz, F. Messina, A. Valencia, A. Funahashi, H. Kitano, C. Auffray, R. Balling, R. Schneider, COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms, Sci. Data 7 (1) (2020) 136.

[7] D. Forni, R. Cagliani, M. Clerici, M. Sironi, Molecular evolution of human coronavirus genomes, Trends Microbiol. 25 (1) (2017) 35–48.

[8] S. Ozono, Y. Zhang, H. Ode, T.S. Tan, K. Imai, K. Miyoshi, S. Kishigami, T. Ueno, Y. Iwatani, T. Suzuki, K. Tokunaga, Naturally mutated spike proteins of SARS-CoV-2 variants show differential levels of cell entry, bioRxiv 2020 (2020), 06.15.151779.

[9] D. Kim, J.Y. Lee, J.S. Yang, J.W. Kim, V.N. Kim, H. Chang, The architecture of SARS-CoV-2 transcriptome, Cell 181 (4) (2020) 914–921 e10.

[10] M.R. Islam, M.N. Hoque, M.S. Rahman, A. Alam, M. Akther, J.A. Puspo, S. Akter, M. Sultana, K.A. Crandall, M.A. Hossain, Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity, Sci. Rep. 10 (1) (2020) 14004.

[11] K. Peng, S. Vucetic, P. Radivojac, C.J. Brown, A.K. Dunker, Z. Obradovic, Optimizing long intrinsic disorder predictors with protein evolutionary information, J. Bioinf. Comput. Biol. 3 (1) (2005) 35–60.

[12] Z.L. Peng, L. Kurgan, Comprehensive comparative assessment of in-silico predictors of disordered regions, Curr. Protein Pept. Sci. 13 (1) (2012) 6–18.

[13] X. Fan, L. Kurgan, Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus, J. Biomol. Struct. Dyn. 32 (3) (2014) 448–464.

[14] F. Meng, V.N. Uversky, L. Kurgan, Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions, Cell. Mol. Life Sci. 74 (17) (2017) 3069–3090.

[15] W. Shen, S. Le, Y. Li, F. Hu, SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation, PLoS One 11 (10) (2016), e0163962.

[16] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (6) (2018) 1547–1549.

[17] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinf. 5 (2004) 113.

[18] M.V. Han, C.M. Zmasek, phyloXML: XML for evolutionary biology and comparative genomics, BMC Bioinf. 10 (2009) 356.

[19] Y. Ren, T. Shu, D. Wu, J. Mu, C. Wang, M. Huang, Y. Han, X.Y. Zhang, W. Zhou, Y. Qiu, X. Zhou, The ORF3a protein of SARS-CoV-2 induces apoptosis in cells, Cell. Mol. Immunol. 17 (8) (2020) 881–883.

[20] S.S. Hassan, P.P. Choudhury, P. Basu, S.S. Jana, Molecular conservation and differential mutation on ORF3a gene in Indian SARS-CoV2 genomes, Genomics 112 (5) (2020) 3226–3237.

[21] E. Issa, G. Merhi, B. Panossian, T. Salloum, S. Tokajian, SARS-CoV-2 and ORF3a: nonsynonymous mutations, functional domains, and viral pathogenesis, mSystems 5 (3) (2020).

[22] A. Shah, Novel coronavirus-induced NLRP3 inflammasome activation: a potential drug target in the treatment of COVID-19, Front. Immunol. 11 (2020) 1021.

[23] Y. Rui, J. Su, S. Shen, Y. Hu, D. Huang, W. Zheng, M. Lou, Y. Shi, M. Wang, S. Chen, N. Zhao, Q. Dong, Y. Cai, R. Xu, S. Zheng, X.F. Yu, Unique and complementary suppression of cGAS-STING and RNA sensing- triggered innate immune responses by SARS-CoV-2 proteins, Sign. Transduct. Target Ther. 6 (1) (2021) 123.

[24] C.N. Cavasotto, M.S. Lamas, J. Maggini, Functional and druggability analysis of the SARS-CoV-2 proteome, Eur. J. Pharmacol. 890 (2021) 173705.

[25] J.Y. Lam, C.K. Yuen, J.D. Ip, W.M. Wong, K.K. To, K.Y. Yuen, K.H. Kok, Loss of orf3b in the circulating SARS-CoV-2 strains, Emerg. Microb. Infect. 9 (1) (2020) 2685–2696.

[26] A. Hachim, H. Gu, O. Kavian, M.Y. Kwan, W.H. Chan, Y.S. Yau, S.S. Chiu, O. T. Tsang, D.S. Hui, F. Ma, E.H. Lau, S.M. Cheng, L.L. Poon, J.M. Peiris, S. A. Valkenburg, N. Kavian, The SARS-CoV-2 antibody landscape is lower in magnitude for structural proteins, diversified for accessory proteins and stable long-term in children, medRxiv (2021), https://doi.org/10.1101/2021.01.03.21249180.

[27] E.H. Mattar, F. Elrashdy, H.A. Almehdar, V.N. Uversky, E.M. Redwan, Natural resources to control COVID-19: could lactoferrin amend SARS-CoV-2 infectivity? PeerJ 9 (2021), e11303.

[28] A. Hachim, N. Kavian, C.A. Cohen, A.W.H. Chin, D.K.W. Chu, C.K.P. Mok, O.T. Y. Tsang, Y.C. Yeung, R. Perera, L.L.M. Poon, J.S.M. Peiris, S.A. Valkenburg, ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection, Nat. Immunol. 21 (10) (2020) 1293–1301.

[29] Y. Konno, I. Kimura, K. Uriu, M. Fukushi, T. Irie, Y. Koyanagi, D. Sauter, R. J. Gifford, S. Nakagawa, K. Sato, SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is increased by a naturally occurring elongation variant, Cell Rep. 32 (12) (2020) 108185.

[30] J.Y. Li, C.H. Liao, Q. Wang, Y.J. Tan, R. Luo, Y. Qiu, X.Y. Ge, The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway, Virus Res. 286 (2020) 198074.

[31] V. Gunalan, A. Mirazimi, Y.J. Tan, 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 (2011) 446.

[32] X. Lei, X. Dong, R. Ma, W. Wang, X. Xiao, Z. Tian, C. Wang, Y. Wang, L. Li, L. Ren, F. Guo, Z. Zhao, Z. Zhou, Z. Xiang, J. Wang, Activation and evasion of type I interferon responses by SARS-CoV-2, Nat. Commun. 11 (1) (2020) 3810.

[33] H. Xia, Z. Cao, X. Xie, X. Zhang, J.Y. Chen, H. Wang, V.D. Menachery, R. Rajsbaum, P.Y. Shi, Evasion of type I interferon by SARS-CoV-2, Cell Rep. 33 (1) (2020) 108234.

[34] L.A. Holland, E.A. Kaelin, R. Maqsood, B. Estifanos, L.I. Wu, A. Varsani, R. U. Halden, B.G. Hogue, M. Scotch, E.S. Lim, An 81-nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (January to March 2020), J. Virol. 94 (14) (2020).

[35] F. Pereira, Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene, Infect. Genet. Evol. 85 (2020) 104525.

[36] S.S. Hassan, A.A.A. Aljabali, P.K. Panda, S. Ghosh, D. Attrish, P.P. Choudhury, M. Seyran, D. Pizzol, P. Adadi, T.M. Abd El-Aziz, A. Soares, R. Kandimalla, K. Lundstrom, A. Lal, G.K. Azad, V.N. Uversky, S.P. Sherchan, W. Baetas-da-Cruz, B.D. Uhal, N. Rezaei, G. Chauhan, D. Barh, E.M. Redwan, G.W. Dayhoff 2nd, N. G. Bazan, A. Serrano-Aroca, A. El-Demerdash, Y.K. Mishra, G. Palu, K. Takayama, A.M. Brufsky, M.M. Tambuwala, A unique view of SARS-CoV-2 through the lens of ORF8 protein, Comput. Biol. Med. 133 (2021) 104380.

[37] Y.C.F. Su, D.E. Anderson, B.E. Young, M. Linster, F. Zhu, J. Jayakumar, Y. Zhuang, S. Kalimuddin, J.G.H. Low, C.W. Tan, W.N. Chia, T.M. Mak, S. Octavia, J. M. Chavatte, R.T.C. Lee, S. Pada, S.Y. Tan, L. Sun, G.Z. Yan, S. Maurer-Stroh, I. H. Mendenhall, Y.S. Leo, D.C. Lye, L.F. Wang, G.J.D. Smith, Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2, mBio 11 (4) (2020).

[38] R. Cagliani, D. Forni, M. Clerici, M. Sironi, Coding potential and sequence conservation of SARS-CoV-2 and related animal viruses, Infect. Genet. Evol. 83 (2020) 104353.

[39] S.S. Hassan, D. Attrish, S. Ghosh, P.P. Choudhury, V.N. Uversky, A.A.A. Aljabali, K. Lundstrom, B.D. Uhal, N. Rezaei, M. Seyran, D. Pizzol, P. Adadi, A. Soares, T. M. Abd El-Aziz, R. Kandimalla, M.M. Tambuwala, G.K. Azad, S.P. Sherchan, W. Baetas-da-Cruz, A. Lal, G. Palu, K. Takayama, A. Serrano-Aroca, D. Barh, A. M. Brufsky, Notable sequence homology of the ORF10 protein introspects the architecture of SARS-CoV-2, Int. J. Biol. Macromol. 181 (2021) 801–809.

[40] F.K. Yoshimoto, The proteins of severe acute respiratory Syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19, Protein J. 39 (3) (2020) 198–216.

[41] J.S. Kim, J.H. Jang, J.M. Kim, Y.S. Chung, C.K. Yoo, M.G. Han, Genome-wide identification and characterization of point mutations in the SARS-CoV-2 genome, Osong Public Health Res. Perspect. 11 (3) (2020) 101–111.

[42] B. Xue, R.L. Dunbrack, R.W. Williams, A.K. Dunker, V.N. Uversky, PONDR-FIT: a meta-predictor of intrinsically disordered amino acids, Biochim. Biophys. Acta Protein Proteonomics 1804 (4) (2010) 996–1010.

[43] P. Romero, Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, A.K. Dunker, Sequence complexity of disordered protein, Proteins: Struct, Funct. Bioinf. 42 (1) (2001) 38–48.

[44] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, A.K. Dunker, Exploiting heterogeneous sequence properties improves prediction of protein disorder, Proteins: Struct, Funct. Bioinf. 61 (S7) (2005) 176–182.

[45] K. Peng, P. Radivojac, S. Vucetic, A.K. Dunker, Z. Obradovic, Length-dependent prediction of protein intrinsic disorder, BMC Bioinf. 7 (1) (2006) 208.

[46] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, C.J. Brown, A.K. Dunker, Predicting intrinsic disorder from amino acid sequence, Proteins 53 (Suppl 6) (2003) 566–572.

[47] J. Prilusky, C.E. Felder, T. Zeev-Ben-Mordehai, E.H. Rydberg, O. Man, J. S. Beckmann, I. Silman, J.L. Sussman, FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded, Bioinformatics 21 (16) (2005) 3435–3438.

[48] Z. Dosztányi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (16) (2005) 3433–3434.
[49] A. Campen, R.M. Williams, C.J. Brown, J. Meng, V.N. Uversky, A.K. Dunker, TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder, Protein Pept. Lett. 15 (9) (2008) 956–963.
[50] M. Seyran, D. Pizzol, P. Adadi, T.M.A. El-Aziz, S.S. Hassan, A. Soares, R. Kandimalla, K. Lundstrom, M. Tambuwala, A.A.A. Aljabali, A. Lal, G.K. Azad, P. P. Choudhury, V.N. Uversky, S.P. Sherchan, B.D. Uhal, N. Rezaei, A.M. Brufsky, Questions concerning the proximal origin of SARS-CoV-2, J. Med. Virol. 93 (3) (2021) 1204–1206.
[51] P. V'Kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for SARS-CoV-2, Nat. Rev. Microbiol. 19 (3) (2021) 155–170.
[52] A. Brufsky, M.T. Lotze, Ratcheting down the virulence of SARS-CoV-2 in the COVID-19 pandemic, J. Med. Virol. 92 (11) (2020) 2379–2380.
[53] P. Majumdar, S. Niyogi, SARS-CoV-2 mutations: the biological trackway towards viral fitness, Epidemiol. Infect. 149 (2021) e110.
[54] M. Bianchi, A. Borsetti, M. Ciccozzi, S. Pascarella, SARS-Cov-2 ORF3a: Mutability and function, Int. J. Biol. Macromol. 170 (2021) 820–826.
[55] S. Wu, C. Tian, P. Liu, D. Guo, W. Zheng, X. Huang, Y. Zhang, L. Liu, Effects of SARS-CoV-2 mutations on protein structures and intraviral protein-protein interactions, J. Med. Virol. 93 (4) (2021) 2132–2140.
[56] M. Miao, E. Clercq, G. Li, Genetic diversity of SARS-CoV-2 over a one-year period of the COVID-19 pandemic: a global perspective, Biomedicines 9 (4) (2021).
[57] D.K.W. Chu, K.P.Y. Hui, H. Gu, R.L.W. Ko, P. Krishnan, D.Y.M. Ng, G.Y.Z. Liu, C.K. C. Wan, M.C. Cheung, K.C. Ng, J.M. Nicholls, D.N.C. Tsang, M. Peiris, M.C. W. Chan, L.L.M. Poon, Introduction of ORF3a-Q57H SARS-CoV-2 variant causing fourth epidemic wave of COVID-19, Hong Kong, China, Emerg. Infect. Dis. 27 (5) (2021) 1492–1495.
[58] D.C. Groves, S.L. Rowland-Jones, A. Angyal, The D614G mutations in the SARS-CoV-2 spike protein: implications for viral infectivity, disease severity and vaccine design, Biochem. Biophys. Res. Commun. 538 (2021) 104–107.
[59] S.S. Hassan, V. Kodakandla, E.M. Redwan, K. Lundstrom, P.P. Choudhury, T. M. Abd El-Aziz, K. Takayama, R. Kandimalla, A. Lal, Á. Serrano-Aroca, G.K. Azad, A.A.A. Aljabali, G. Palu, G. Chauhan, P. Adadi, M. Tambuwala, A.M. Brufsky, W. Baetas-da-Cruz, D. Barh, N.G. Bazan, V.N. Uversky, An issue of concern: unique truncated ORF8 protein variants of SARS-CoV-2, bioRxiv 5 (25) (2021) 445557, 2021.