Variability of Accessory Proteins Rules the SARS-CoV-2 Pathogenicity

Sk. Sarif Hassan[a,*], Pabitra Pal Choudhury[b], Vladimir N. Uversky[c,*], Guy W. Dayhoff II[d], Alaa A. A. Aljabali[e], Bruce D. Uhal[f], Kenneth Lundstrom[g], Nima Rezaei[h], Murat Seyran[i], Damiano Pizzol[j], Parise Adadi[k], Amos Lal[l], Antonio Soares[m], Tarek Mohamed Abd El-Aziz[n], Ramesh Kandimalla[o], Murtaza Tambuwala[p], Gajendra Kumar Azad[q], Samendra P. Sherchan[r], Wagner Baetas-da-Cruz[s], Kazuo Takayama[t], Ángel Serrano-Aroca[u], Gaurav Chauhan[v], Giorgio Palu[w], Adam M. Brufsky[x]

[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 Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA*
[d] *Department of Chemistry, College of Art and Sciences, University of South Florida, Tampa, FL 33620, USA*
[e] *Department of Pharmaceutics and Pharmaceutical Technology, Yarmouk University-Faculty of Pharmacy, Irbid 566, Jordan*
[f] *Department of Physiology, Michigan State University, East Lansing, MI 48824, USA*
[g] *PanTherapeutics, Rte de Lavaux 49, CH1095 Lutry, Switzerland*
[h] *Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran &*
Network of Immunity in Infection, Malignancy and Autoimmunity (NIIMA), Universal Scientific Education and Research Network (USERN), Stockholm, Sweden
[i] *Doctoral studies in natural and technical sciences (SPL 44), University of Vienna, Austria*
[j] *Italian Agency for Development Cooperation - Khartoum, Sudan Street 33, Al Amarat, Sudan*
[k] *Department of Food Science, University of Otago, Dunedin 9054, New Zealand*
[l] *Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, Minnesota, USA*
[m] *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*
[n] *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 &*
Zoology Department, Faculty of Science, Minia University, El-Minia 61519, Egypt
[o] *Applied Biology, CSIR-Indian Institute of Chemical Technology Uppal Road, Tarnaka, Hyderabad-500007, Telangana State, India &*
Department of Biochemistry, Kakatiya Medical College, Warangal, Telangana, India
[p] *School of Pharmacy and Pharmaceutical Science, Ulster University, Coleraine BT52 1SA, Northern Ireland, UK*
[q] *Department of Zoology, Patna University, Patna-800005, Bihar, India*
[r] *Department of Environmental Health Sciences, Tulane University, New Orleans, LA, 70112, USA*
[s] *Translational Laboratory in Molecular Physiology, Centre for Experimental Surgery, College of Medicine, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil*
[t] *Center for iPS Cell Research and Application, Kyoto University, Japan*
[u] *Biomaterials and Bioengineering Lab, Translational Research Centre San Alberto Magno, Catholic University of Valencia San Vicente Mártir, c/Guillem de Castro 94, 46001 Valencia, Spain*
[v] *School of Engineering and Sciences, Tecnologico de Monterrey, Av. Eugenio Garza Sada 2501 Sur, 64849 Monterrey, Nuevo León, Mexico*
[w] *Department of Molecular Medicine, University of Padova, Via Gabelli 63, 35121, Padova, Italy*
[x] *University of Pittsburgh School of Medicine, Department of Medicine, Division of Hematology/Oncology, UPMC Hillman Cancer Center, Pittsburgh, PA, USA*

Abstract

The coronavirus disease 2019 (COVID-19) is caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) which is pandemic with an estimated fatality rate less than 1% is ongoing. SARS-CoV-2 accessory proteins ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 with putative functions to manipulate host immune mechanisms such as interferons, immune signaling receptor NLRP3 (NOD-, LRR-, and pyrin domain-containing 3) inflammasome, inflammatory cytokines such as interleukin 1β (IL-1β) are critical in COVID-19 pathology. Outspread variations of each of the six accessory proteins of all complete proteomes (available as of October 26, 2020, in the National Center for Biotechnology Information

*Corresponding author
Email addresses: sarimif@gmail.com (Sk. Sarif Hassan), pabitrapalchoudhury@gmail.com (Pabitra Pal Choudhury), vuversky@usf.edu (Vladimir N. Uversky), gdayhoff@usf.edu (Guy W. Dayhoff II), alaaj@yu.edu.jo (Alaa A. A. Aljabali), bduhal@gmail.com (Bruce D. Uhal), lundstromkenneth@gmail.com (Kenneth Lundstrom), rezaei_nima@tums.ac.ir (Nima Rezaei), muratseyran@gmail.com (Murat Seyran), damianopizzol8@gmail.com (Damiano Pizzol), parise.adadi@postgrad.otago.ac.nz (Parise Adadi), manavamos@gmail.com (Amos Lal), soaresa@uthscsa.edu (Antonio Soares), mohamedt1@uthscsa.edu (Tarek Mohamed Abd El-Aziz), ramesh.kandimalla@iict.res.in (Ramesh Kandimalla), m.tambuwala@ulster.ac.uk (Murtaza Tambuwala), gkazad@patnauniversity.ac.in (Gajendra Kumar Azad), sshercha@tulane.edu (Samendra P. Sherchan), wagner.baetas@gmail.com (Wagner Baetas-da-Cruz), kazuo.takayama@cira.kyoto-u.ac.jp (Kazuo Takayama), angel.serrano@ucv.es (Ángel Serrano-Aroca), gchauhan@tec.mx (Gaurav Chauhan), giorgio.palu@unipd.it (Giorgio Palu), brufskyam@upmc.edu (Adam M. Brufsky)

depository) of SARS-CoV-2, were observed across six continents. Across all continents, the decreasing order of percentage of unique variations in the accessory proteins was found to be ORF3a>ORF8>ORF7a>ORF6>ORF10>ORF7b. The highest and lowest unique variations of ORF3a were observed in South America and Oceania, respectively. This finding suggests that the wide variations of accessory proteins seem to govern the pathogenicity of SARS-CoV-2, and consequently, certain propositions and recommendations can be made in the public interest.

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

1. Introduction

SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2), the causative agent for The coronavirus disease 2019 (COVID-19), the pandemic is ongoing with the estimated fatality rate less than 1% [1]. However, the World Health Organization (WHO), Health Emergencies Programme, Executive Director, Dr Michael Ryan, in 2020 October indicated that 760 million might have been infected with SARS-CoV-2 which makes the hypothetical fatality rate as 0.14% with approximately one million lives taken. SARS-CoV-2 is a member of Betacoronavirus (lineage B) genus and Sarbecovirus subgenus was suggested to diverge from the lineage of Bat Coronavirus (BatCoV) Ratg13 in 1969 with the 95% highest posterior density interval of the years 1930 to 2000 [2]. Amongst previous Human Coronaviruses (HuCoVs) Severe acute respiratory syndrome-related coronavirus (SARS-CoV) is the closest to SARS-CoV-2 that caused an endemic from 2002 to 2004 [2, 3]. SARS-CoV had 8 open reading frame (ORF) 3a, 3b, 6, 7a, 7b, 8a, 8b, and 9b which were suggested to have more intrinsic and secondary other than having primary roles in the cell cycle and cellular entry [4, 5]. For instance, the ORFs are transcripted through the second phase of replication by (+) subgenomic messenger RNAs that were transcripted by the viral replication transcription complex negative-sense viral RNA coded in the initial stages of SARS-CoV-2 infection [6]. Thus, due to their intrinsic nature accessory proteins are not positive-selection sites such as extrinsic and primary functional Spike protein receptor-binding domain or protease cleavage sites [7]. Since, clinical SARS-CoV-2 isolates had a high-frequency non-synonymous mutation, D614G, in their S protein, which increased host cell entry via ACE2 and Transmembrane Protease Serine 2 (TMPRSS2) [8]. Therefore, due to the intrinsic nature and secondary order in the viral transcription, we can expect less selective pressure and mutations on accessory proteins to reach high-frequency in the population is less expected. Thus, despite the 19 to 89 years of estimated genomic divergence between SARS-CoV-2 and Ratg13, the sequence identity on accessory proteins ORF3, ORF6, ORF7a, ORF7b, ORF8, ORF10 had 98.5, 100, 97.5, 97.6, 95, and 100%, respectively are very high or identical which is indicating somehow the direct ancestor of SARS-CoV-2 had been exposed to almost no selective pressure to manipulate its intermediate host immunity for many years until the primary Human infection in Wuhan (Fig.1 to Fig.6) [2].

SARS-CoV-2 and SARS-CoV accessory proteins have differences such as putative protein ORF10 not present in SARS-CoV and the ORF3b and ORF9b is not present in SARS-CoV-2 [9, 10]. Very little is known about the functions of accessory proteins of SARS-CoV-2. The known essential features of the six accessory proteins are summarized below.

ORF3a Protein: The ORF3a is the 275 amino acids long largest accessory protein among the accessory proteins coded by the SARS-CoV-2, has 72.4% sequence identity with SARS-CoV ORF3a protein and has 98.5% sequence identity with BatCoV Ratg13 ORF3a protein [11, 12] (Fig.1).

```
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
                ******************.****************
```

Figure 1: ClustalW alignment of SARS-CoV-2 and Ratg13 ORF3 proteins shows 98.5% sequence identity.

ORF3a is involved in virulence, infectivity, ion channel activity, morphogenesis, and virus release [13]. In SARS-CoV ORF3a was a multifunctional protein, co-localized with its protein binding regions with E, M, and S proteins in viral assembly formed homo-tetrameric complex as potassium-ion channel on host membrane [5]. In SARS-CoV-2, the ion-channel proteins (viroporins) function of ORF3a in addition to other proteins such as protein E, and ORF8a is critical in CoVs tissue inflammation [6]. Viroporins mediated lysosomal disruption and ion-redistribution activates innate immune signaling receptor NLRP3 (NOD-, LRR-, and 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 with its protein binding domains interacts with TNF receptor-associated factor (TRAF3) protein, which leads ASC ubiquitination, and caspase 1 activation, and IL-1β maturation [14]. Additionally, ORF3a and ORF7a in combination with E, S, Nsp1 protein, and MAPK pathway proteins (MAPK8, MAPK14, and MAP3K7) trigger proinflammatory cytokine signaling transcription factors such as STAT1, STAT2, IRF9, and NFKB1 [6]. Another SARS-CoV-2 ORF3a protein interacts with heme oxygenase-1 (HMOX1) that has a role in heme catabolism and the anti-inflammatory system [6]. SARS-CoV-2 either triggers viral dissemination or suppresses continued viral replication of the apoptosis or programmed cell death [6]. In SARS-CoV ORF3a E and M protein, ion channel activity interferes with apoptotic pathways [11].

ORF6 Protein: SARS-CoV-2 ORF6 is a 61 amino acid long membrane-associated interferon (IFN) antagonist protein suppresses the expression of co-transfected expression constructs and its subcellular localization to vesicular structures that has 68.9% sequence identity, with SARS-CoV ORF6 protein and has 100% sequence identity, with BatCoV Ratg13 ORF6 protein[5] (Fig.2).

```
BCA87365.1      MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKNLSKSLTENKYSQLDEEQPMEI
MN996532.2      MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKNLSKSLTENKYSQLDEEQPMEI
                ************************************************************

BCA87365.1      D
MN996532.2      D
                *
```

Figure 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 shows 100% sequence identity, despite up to 89 years of genetic diversion.

ORF6 interacts with the karyopherin import complex that limits the movement of transcription factors STAT1 which down-regulates the IFN pathway [5]. In SARS-CoV ORF6 and ORF3a, in association with other proteins such as M, Nsp1 with Nsp3 inhibit IRF3 signalling and repress interferon expression and stimulate the degradation of IFNAR1 and STAT1 [6]. ORF6 interacts with the nsp8 protein coded by SARS-CoV-2 and it can increase infection during early infection at a low multiplicity with increase in RNA polymerase activity [15]. It is reported that ORF6 and ORF8 can inhibit the type-I interferon signaling pathway [15]. ORF6 protein with lysosomal targeting motif (YSEL) and diacidic motif (DDEE) induces intracellular membrane rearrangements resulting in a vesicular population and endosomal internalization of viral protein into the infected cells, increasing replication [16].

ORF7a and ORF7b Proteins:

ORF7a 121 aa coding type I transmembrane protein interacts with SARS-CoV-2 structural proteins M, E, and S, which are essential for viral assembly, and hence ORF7a is involved in the viral replication cycle and virion-associated ORF7a protein may function during early infection that has 85.2% Sequence identity with SARS-CoV ORF7a protein and has 97.5% sequence identity, with BatCoV Ratg13 ORF7a protein[5] (Fig.3).

```
BCA87366.1      MKIILFLALITLATCELYHYQECVRGTTVLLKEPCSSGTYEGNSPFHPLADNKFALTCFS
MN996532.2      MKIILFLVLVTLATCELYHYQECVRGTTVLLKEPCSSGTYEGNSPFHPLADNKFALTCFS
                *******.*:**************************************************

BCA87366.1      TQFAFACPDGVKHVYQLRARSVSPKLFIRQEEVQELYSPIFLIVAAIVFITLCFTLKRKT
MN996532.2      TQFAFACPDGVKHVYQLRARSVSPKLFIRQEEVQELYSPIFLIIAAIVFITLCFTLKRKT
                ******************************************:***************

BCA87366.1      E
MN996532.2      E
                *
```

Figure 3: ClustalW alignment of SARS-CoV-2 (NCBI GenBank ID BCA87366.1) and Ratg13 (NCBI GenBank ID MN996532.2, translated 5'3' frame 2) ORF7a proteins shows 97.5% sequence identity, despite up to 89 years of genetic diversion.

ORF7a interacts with SARS-CoV-2 structural proteins: membrane (M), envelope (E), and spike (S), which are essential for viral assembly, and hence ORF7a is involved in the viral replication cycle and virion-associated ORF7a protein may

function during early infection [17, 18]. ORF7a leads the activation of pro-inflammatory cytokines and chemokines, such as IL-8 and RANTES [5]. SARS-CoV ORF7a in combination with E protein activate apoptosis by suppressing anti-apoptotic protein [6]. ORF7b is a 43 aa coding protein found in association with intracellular virus particles and also in purified virions inside the Golgi compartment that has an 85.4% sequence identity with SARS-CoV ORF7b protein and has 97.6% sequence identity, with BatCoV Ratg13 ORF7a protein [5] (Fig.4).

```
BCB15096.1      MIELSLIDFYLCFLAFLLFLVLIMLIIFWFSLELQDHNETCHA
MN996532.2      MSELSLIDFYLCFLAFLLFLVLIMLIIFWFSLELQDHNETCHA
                * ****************************************
```

Figure 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.

ORF7b is found in association with intracellular virus particles and also in purified virions. Till date, there is very little experimental evidence to support a role for ORF7a or ORF7b in the replication of SARS-CoV-2 [19].

ORF8 Protein: ORF8 (121 aa long) is a unique accessory protein in SARS-CoV-2, and it stands out by being poorly conserved among other CoVs and accordingly showing structural changes suggested to be related to the ability of the virus to spread [20]. ORF8 is a unique accessory protein in SARS-CoV-2, which stands out by being poorly conserved among other CoVs and accordingly showing structural changes suggested to be related to the ability of the virus to spread [20]. ORF8 sequences of SARS-CoV-2 and Ratg13 share 95% amino acid identity (Fig.5).

```
QJA17759.1      MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSKWYIRVGARKSAPLIEL
MN996532.2      MKLLVFLGILTTVTAFHQECSLQSCAQHQPYVVDDPCPIHFYSKWYIRVGARKSAPLIEL
                **:******:***:***********:**********************************

QJA17759.1      CVDEAGSKSPIQYIDIGNYTVSCLPFTINCQEPKLGSLVVRCSFYEDFLEYHDVRVVLDF
MN996532.2      CVDEVGSKSPIQYIDIGNYTVSCSPFTINCQEPKLGSLVVRCSFYEDFLEYHDVRVVLDF
                ****.***************** ***********************************

QJA17759.1      I
MN996532.2      I
                *
```

Figure 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 shows 95% sequence identity, despite up to 89 years of genetic diversion.

ORF8 of SARS-CoV-2 interacts with major histocompatibility complex (MHC) class-I molecules and down-regulates their surface expression significantly on various cell types [21]. It has been reported earlier that inhibition of ORF8 function could be a strategy to improve the special immune surveillance and accelerate the eradication of SARS-CoV-2 *in vivo* [22].

ORF10 Protein: The accessory protein 38 aa coding protein ORF10 has been reported to be unique for SARS-CoV-2 containing eleven cytotoxic T lymphocyte (CTL) epitopes of nine amino acids in length each, across various human leukocyte antigen (HLA) subtypes [23, 24]. ORF10 negatively affects the antiviral protein degradation process through its interaction

with the Cul2 ubiquitin ligase complex [6]. SARS-CoV does not have ORF10 protein but SARS-COV-2 ORF10 and Ratg13 ORF10 has 97.3% sequence identity [25] (Fig.6).

```
BCA87369.1      MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNFNLT
MN996532.2      MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNLNLT
                ************************************:***
```

Figure 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 shows 97.3% sequence identity, despite up to 89 years of genetic diversion.

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

2. Data acquisition

Sequences for all the accessory proteins ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 were downloaded (on October, 20, 2020) from the complete SARS-CoV-2 proteomes on the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/) (Table 1).

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. 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* (7-13). There were 510, 72, 158, 37, 190, and 44 unique accessory proteins ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10, respectively, available. For each continent, ranges and names of sequences are presented in Table 2.

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

2.1. Evaluating 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 [26, 27, 28, 29]. The per-residue disorder predisposition scores are on a scale from 0 to 1, where values of 0 indicate fully ordered residues, and values of 1 indicate 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 taken as moderately flexible.

3. Results

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 accessory proteins ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 sequences are 419, 55, 122, 26, 147, and 32, respectively (Supplementary Fig.11). Furthermore, the percentage of unique sequences on each continent among all available accessory proteins are also enumerated (Fig.7).

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

Figure 7: Number of unique accessory proteins across six continents

The percentage of each accessory protein across the six continents are presented as bar diagrams in Fig.8.

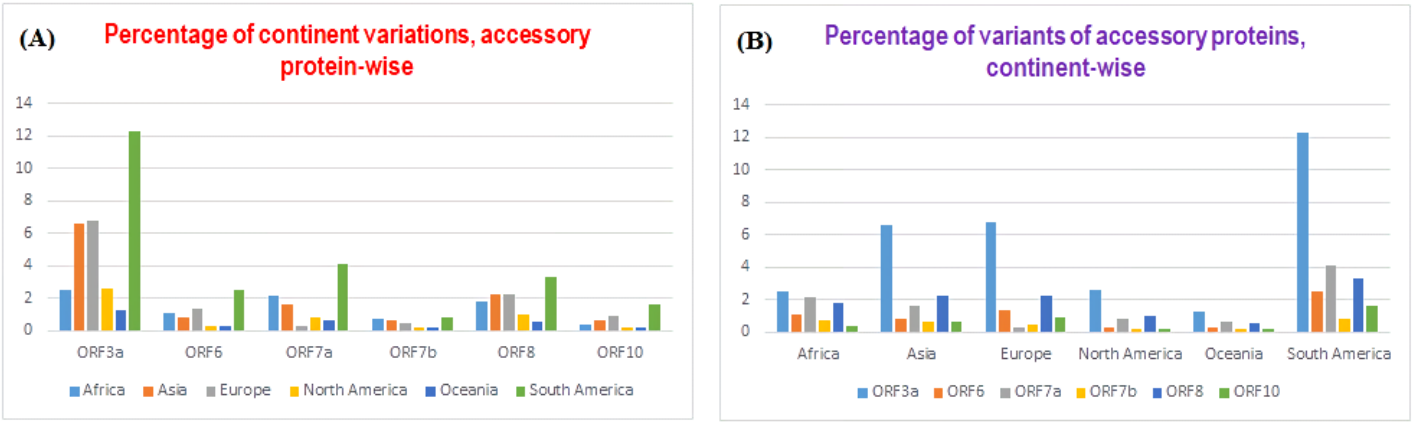


Figure 8: Bar representations of percentages of continental variations (A), and the percentage of unique accessory proteins (B).

From the Fig.8, the following observations were drawn:

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 least number of unique variations in 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 was 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 amount (significantly) of unique variations across all the 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.

In addition, the percentage of unique accessory proteins among all unique sequences obtained across the six continents are represented as bar diagrams in Fig.9.

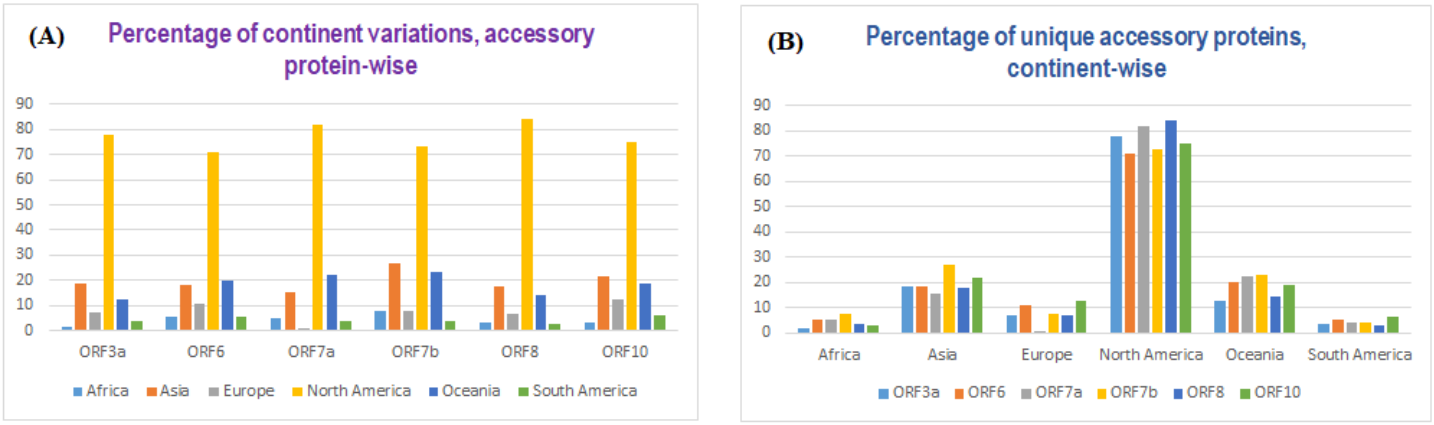
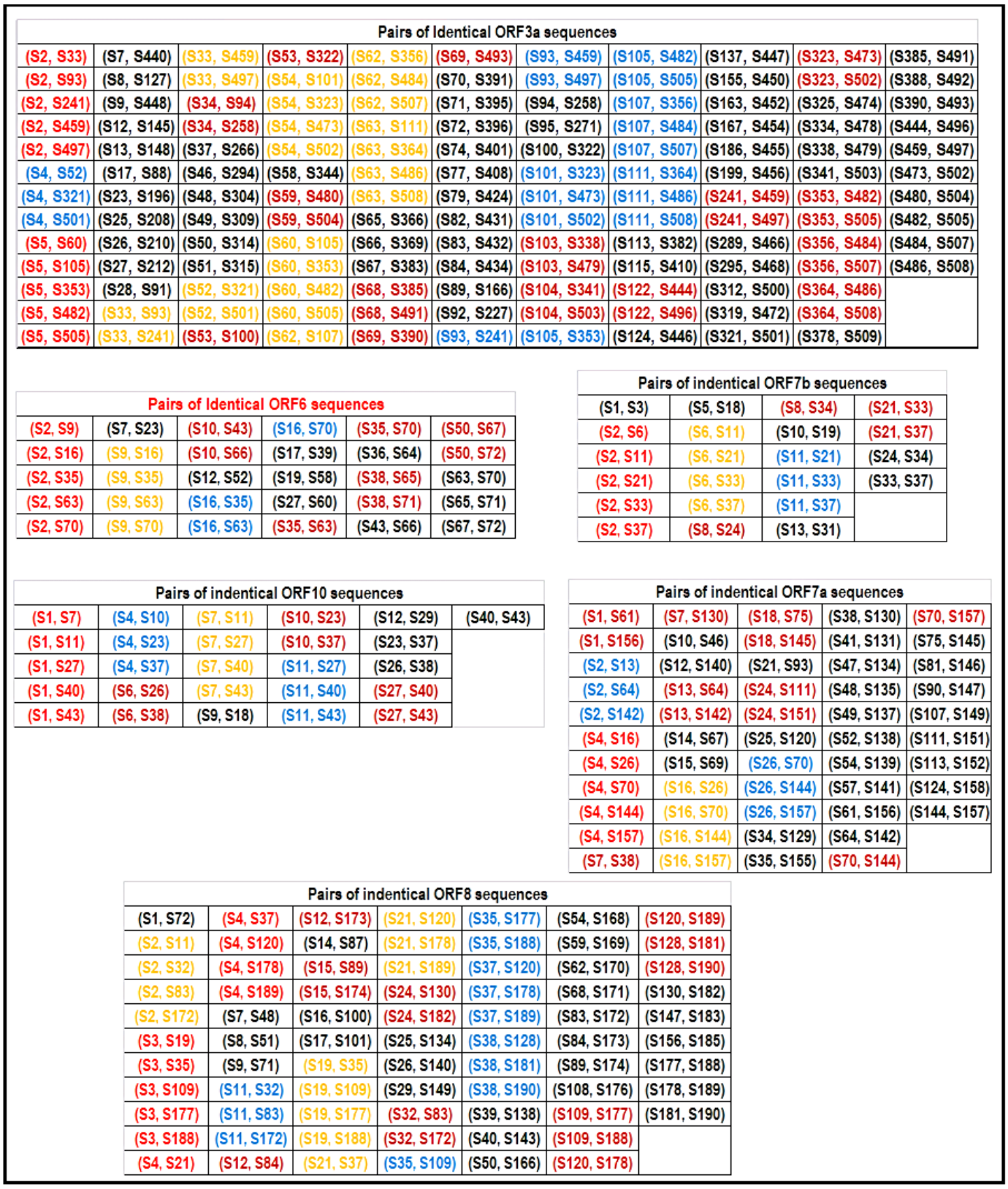


Figure 9: Quantitative information of the accessory proteins

The percentages of each unique variations among all unique variants of all accessory proteins across all six continents can be concluded from Fig.9 as follows:

Among all available unique variation of the six accessory proteins of SARS-CoV-2, North America and South America exhibited the highest and lowest percentage of variations of each accessory protein, 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. With regards to 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 varieties of ORF8 and ORF7a (among all accessory proteins) was found in North America and Europe, respectively.

Following continent-wise, lists of identical sequences for each accessory protein were presented (Fig.10).



Pairs of Identical ORF3a sequences

(S2, S33)	(S7, S440)	(S33, S459)	(S53, S322)	(S62, S356)	(S69, S493)	(S93, S459)	(S105, S482)	(S137, S447)	(S323, S473)	(S385, S491)
(S2, S93)	(S8, S127)	(S33, S497)	(S54, S101)	(S62, S484)	(S70, S391)	(S93, S497)	(S105, S505)	(S155, S450)	(S323, S502)	(S388, S492)
(S2, S241)	(S9, S448)	(S34, S94)	(S54, S323)	(S62, S507)	(S71, S395)	(S94, S258)	(S107, S356)	(S163, S452)	(S325, S474)	(S390, S493)
(S2, S459)	(S12, S145)	(S34, S258)	(S54, S473)	(S63, S111)	(S72, S396)	(S95, S271)	(S107, S484)	(S167, S454)	(S334, S478)	(S444, S496)
(S2, S497)	(S13, S148)	(S37, S266)	(S54, S502)	(S63, S364)	(S74, S401)	(S100, S322)	(S107, S507)	(S186, S455)	(S338, S479)	(S459, S497)
(S4, S52)	(S17, S88)	(S46, S294)	(S58, S344)	(S63, S486)	(S77, S408)	(S101, S323)	(S111, S364)	(S199, S456)	(S341, S503)	(S473, S502)
(S4, S321)	(S23, S196)	(S48, S304)	(S59, S480)	(S63, S508)	(S79, S424)	(S101, S473)	(S111, S486)	(S241, S459)	(S353, S482)	(S480, S504)
(S4, S501)	(S25, S208)	(S49, S309)	(S59, S504)	(S65, S366)	(S82, S431)	(S101, S502)	(S111, S508)	(S241, S497)	(S353, S505)	(S482, S505)
(S5, S60)	(S26, S210)	(S50, S314)	(S60, S105)	(S66, S369)	(S83, S432)	(S103, S338)	(S113, S382)	(S289, S466)	(S356, S484)	(S484, S507)
(S5, S105)	(S27, S212)	(S51, S315)	(S60, S353)	(S67, S383)	(S84, S434)	(S103, S479)	(S115, S410)	(S295, S468)	(S356, S507)	(S486, S508)
(S5, S353)	(S28, S91)	(S52, S321)	(S60, S482)	(S68, S385)	(S89, S166)	(S104, S341)	(S122, S444)	(S312, S500)	(S364, S486)	
(S5, S482)	(S33, S93)	(S52, S501)	(S60, S505)	(S68, S491)	(S92, S227)	(S104, S503)	(S122, S496)	(S319, S472)	(S364, S508)	
(S5, S505)	(S33, S241)	(S53, S100)	(S62, S107)	(S69, S390)	(S93, S241)	(S105, S353)	(S124, S446)	(S321, S501)	(S378, S509)	

Pairs of Identical ORF6 sequences

(S2, S9)	(S7, S23)	(S10, S43)	(S16, S70)	(S35, S70)	(S50, S67)
(S2, S16)	(S9, S16)	(S10, S66)	(S17, S39)	(S36, S64)	(S50, S72)
(S2, S35)	(S9, S35)	(S12, S52)	(S19, S58)	(S38, S65)	(S63, S70)
(S2, S63)	(S9, S63)	(S16, S35)	(S27, S60)	(S38, S71)	(S65, S71)
(S2, S70)	(S9, S70)	(S16, S63)	(S35, S63)	(S43, S66)	(S67, S72)

Pairs of indentical ORF7b sequences

(S1, S3)	(S5, S18)	(S8, S34)	(S21, S33)
(S2, S6)	(S6, S11)	(S10, S19)	(S21, S37)
(S2, S11)	(S6, S21)	(S11, S21)	(S24, S34)
(S2, S21)	(S6, S33)	(S11, S33)	(S33, S37)
(S2, S33)	(S6, S37)	(S11, S37)	
(S2, S37)	(S8, S24)	(S13, S31)	

Pairs of indentical ORF10 sequences

(S1, S7)	(S4, S10)	(S7, S11)	(S10, S23)	(S12, S29)	(S40, S43)
(S1, S11)	(S4, S23)	(S7, S27)	(S10, S37)	(S23, S37)	
(S1, S27)	(S4, S37)	(S7, S40)	(S11, S27)	(S26, S38)	
(S1, S40)	(S6, S26)	(S7, S43)	(S11, S40)	(S27, S40)	
(S1, S43)	(S6, S38)	(S9, S18)	(S11, S43)	(S27, S43)	

Pairs of indentical ORF7a sequences

(S1, S61)	(S7, S130)	(S18, S75)	(S38, S130)	(S70, S157)
(S1, S156)	(S10, S46)	(S18, S145)	(S41, S131)	(S75, S145)
(S2, S13)	(S12, S140)	(S21, S93)	(S47, S134)	(S81, S146)
(S2, S64)	(S13, S64)	(S24, S111)	(S48, S135)	(S90, S147)
(S2, S142)	(S13, S142)	(S24, S151)	(S49, S137)	(S107, S149)
(S4, S16)	(S14, S67)	(S25, S120)	(S52, S138)	(S111, S151)
(S4, S26)	(S15, S69)	(S26, S70)	(S54, S139)	(S113, S152)
(S4, S70)	(S16, S26)	(S26, S144)	(S57, S141)	(S124, S158)
(S4, S144)	(S16, S70)	(S26, S157)	(S61, S156)	(S144, S157)
(S4, S157)	(S16, S144)	(S34, S129)	(S64, S142)	
(S7, S38)	(S16, S157)	(S35, S155)	(S70, S144)	

Pairs of indentical ORF8 sequences

(S1, S72)	(S4, S37)	(S12, S173)	(S21, S120)	(S35, S177)	(S54, S168)	(S120, S189)
(S2, S11)	(S4, S120)	(S14, S87)	(S21, S178)	(S35, S188)	(S59, S169)	(S128, S181)
(S2, S32)	(S4, S178)	(S15, S89)	(S21, S189)	(S37, S120)	(S62, S170)	(S128, S190)
(S2, S83)	(S4, S189)	(S15, S174)	(S24, S130)	(S37, S178)	(S68, S171)	(S130, S182)
(S2, S172)	(S7, S48)	(S16, S100)	(S24, S182)	(S37, S189)	(S83, S172)	(S147, S183)
(S3, S19)	(S8, S51)	(S17, S101)	(S25, S134)	(S38, S128)	(S84, S173)	(S156, S185)
(S3, S35)	(S9, S71)	(S19, S35)	(S26, S140)	(S38, S181)	(S89, S174)	(S177, S188)
(S3, S109)	(S11, S32)	(S19, S109)	(S29, S149)	(S38, S190)	(S108, S176)	(S178, S189)
(S3, S177)	(S11, S83)	(S19, S177)	(S32, S83)	(S39, S138)	(S109, S177)	(S181, S190)
(S3, S188)	(S11, S172)	(S19, S188)	(S32, S172)	(S40, S143)	(S109, S188)	
(S4, S21)	(S12, S84)	(S21, S37)	(S35, S109)	(S50, S166)	(S120, S178)	

(Si,Sj) denotes its presence in all continents	(Si,Sj) denotes its presence in four continents
(Si,Sj) denotes its presence in five continents	(Si,Sj) denotes its presence in three continents
(Si,Sj) denotes its presence in two continents	

Figure 10: Identical pair of accessory protein sequences across all the continents.

The following observations were made for each accessory protein based on 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 font), S2 (with reference to Africa, QOI60359) and S5 (with reference to 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, which was a strain determining mutation [30]. It is 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 with respect to 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. Besides, 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 has been found in Asia, Europe and North America only. The deletion mutation V255 occurred in S59 (Asia), which was found in Asia, Oceania, and South America. S68 (Asia) and S69 (Asia) possessed two mutations, H93Y and K67N, 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 noticed only in Europe, North America, and South America. The V13L mutation was found in the 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:

Table 3: List of ORF3a sequences and their presence on two continents only

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	K66N	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	K16N, Q57H	Asia and North America	S321	S171L	North America and South America
S89	V55F	Europe and North America	S325	S177I	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	K67N	North America and Oceania
S155	L46F	North America and Oceania	S444	V13L	Oceania and South America
S163	L53F	North America and Oceania			

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 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, which were found on three continents, North America, Oceania, and South America. The ORF6 unique variant S7 (ORF6, Asia) possesses the E13D mutation, which was found only in Asia and North America. The ORF6 sequence S12 (ORF6, Asia) possessed a set of deletions "FKVSIWNLD" (22-30 aa) and it appeared in Asia and North America only. 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, sequence S36 (ORF6, North America) with the D61H mutation was present in North America and Oceania only.

ORF7a: Mutations are based on the Wuhan ORF7a sequence (YP_009724395).

The Wuhan ORF7a sequence YP_009724395 was found on all continents. V104F was found in the sequence 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 sequence S7 (ORF7a, Asia) in Asia, North America, and Oceania. 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:

Table 4: List of ORF7a sequences and their presence on two continents only

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			

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 the six continents. It is 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 and it was present in Africa and Asia. Sequence S5 (ORF7b, Asia) had the mutation S31L and this sequences was found on the two continents, Asia and North America only. L32F occurred in the sequence S10 (ORF7b, Europe) which was present on the continents Europe and North America. Furthermore, ORF7b Sequence S13 had the mutation L4F and this sequence was found on North America and Oceania.

ORF8: Mutations described below are determined with reference to the Wuhan ORF7b sequence (YP_009724396).

It is observed that, the Wuhan ORF8 YP_009724396 sequence was found on every continent. Also, there was another sequence which is also present in every continent, having the single mutation L84S. The V62L, a single mutation was observed in the sequence S2 (ORF8, Africa) which was found on all continents except South America, whereas ORF8 sequence S38 (Europe) possessed the single mutation A65S and the sequence was found in North America, Oceania, and South America. Further, V62L and L84S two mutations were observed in S12 (ORF8, Asia) and S12 appeared in Asia, North America, and Oceania. Sequence S15 (ORF8, Asia) got a mutation S67F and it was found in Asia, North America, and Oceania. ORF8

sequence S24 (Asia) possessed a single mutation A65V and the sequence was found in Asia, North America, and Oceania.

Table 5: List of ORF8 sequences and their presence on two continents only

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	T12N	Asia and North America	S54	S21N	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	D63N	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			

ORF10: Here, mutations are based on the Wuhan ORF10 sequence (YP_009725255).

The Wuhan ORF10 (YP_009725255) became identical with S1 (ORF10, Africa) and it was found on every continent. ORF10 sequence S6 (ORF10, Asia) had the mutation L37F and the sequence was present on North America, and Oceania only. The only mutation V30L was found in 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. Also, the mutation D31Y appeared in S12 (ORF10, Europe) which was found in Europe and North America only.

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 intrinsic protein disordered residues (IPD), molecular weight (MW), and isoelectric point (IP) of all the unique variants of all accessory proteins (Table 6). The zigzag behavior of the plots of IPD, MW, and IP depicts the wide variability of each accessory protein-variant (Supplementary Fig.12-Fig.15).

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

IPD of unique accessory proteins of SARS-CoV-2					
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
IP 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

The following observations were made based on Table 6.

The amount total dispersions (based on range) of the percentage of IPD and MW of ORF6 variants turned out to be highest whereas the highest amount of total dispersion of IP was observed for ORF10. The smallest amount of total dispersions of the percentage of IPD, MW, and IP were found for ORF3a, ORF10, and ORF7b, respectively. The large value of range and variance of the MW of the unique ORF3a, ORF7a, ORF8, and ORF10 variants imply the wide variability of each set of ORF3a, ORF7a, ORF8, and ORF10 though range and variance of IPD and IP were not much widely spread. In case of unique variance of ORF6, the range and variance of MW and percentage of IPD were found to be large which implied the wide quantitative differences among the unique ORF6 variants. Furthermore, moderately high range and variance associated with the percentage of IPD and MW of ORF7a variants imply its moderate variability.

In line with the previously reported data, Fig.14 and Table 6 show that all SARS-CoV-2 accessory proteins contain different levels of intrinsic disorder. Furthermore, this analysis revealed that intrinsic disorder predispositions can 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 0.25 and 0.5). This is an important observation suggesting that natural variability of SARS-CoV-2 accessory proteins is shaping their structural flexibility.

4. Discussion

SARS-CoV-2 is the first HuCoVs with pandemic capacity due to its highly contagious nature deriving from the structural differences in its spike protein such as flat sialic acid binding domain, tight binding to its entry ACE2 receptor and capacity to be cleaved by furin protease [31]. However, based on the estimated infection number close to one billion by WHO, SARS-CoV-2 highly contagious but relatively a weak viral pathogen considering the overall of infection number has severe infections associated with the multiple organ dysfunctions [6]. This relatively weak pathological feature of SARS-CoV-2 could be related to the accessory proteins modulating host immunity as described above.

Based on the dynamic and various mutations on accessory protein variants, SARS-CoV-2 after diverging with BatCoV Ratg13 19 to 89 years ago was likely to have very few infections or somehow had very few selective pressure to tackle host immunity in nature [2]. In SARS-CoV-2 as other CoVs the genomic stability of their relatively large RNA genome around 30,000 amino acids is protected with proofreading proteins majorly 3'-5' exonuclease non-structural protein (nsp)14 and 205 cofactor proteins nsp10, nsp13, and nsp16 [32]. Muller's ratchet or also called as rachet effect explains the extinctive effect of high mutation rates of asexual organisms such as viruses [33]. Therefore, SARS-CoV-2 is repairing its mutations to preserve its genomic stability since a mutation can lead to pathological fitness losses or viral extinction [33]. 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 in zoonotic hosts [33]. For instance, SARS-CoV mutations, a 29-nucleotide deletion of ORF8, was associated with the less pathogenic strain [33]. Similarly, SARS-CoV-2 variants with a 382-nucleotide deletion, ORF8 had mild symptoms and did not require supplemental oxygen [33].

Furthermore, only one variant (identical to the Wuhan sequence (NC_045512) of each of the accessory proteins of ORF6, ORF7a, ORF7b, and ORF10 were present on all continents. Furthermore, it was observed that only two variants of ORF3a differed by a single mutation (Q57H, a clade/strain determining [30]) were found on all six continents. 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 has turned out to be the smallest. These findings confirmed that the other variants of all the accessory proteins were due to demographic, environmental constraints.

It was found that most of the unique variants of accessory proteins differed from the respective Wuhan accessory proteins by a single mutation, although basic descriptive statistics as found in section 3.1, unfolded their respective wide variability. Note that new variants of each accessory protein have been found in recent days and continue to do so. Significant amounts of unique variants of each accessory protein having wide variability might contribute significantly to the pathogenicity of SARS-CoV-2.

Therefore, our firm conviction that natural weakened stability (if achievable) of SARS-CoV-2 seems to be a far reachable destiny that alarms the danger of the present pandemic scenario due to COVID-19. Also, unique accessory protein variants across individual continents would all be expected to be mixed, while international travels would be restarted without strict protective measures. 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 administrations to take necessary measures to mitigate the spread of COVID-19.

Author Contributions

SSH conceived the project and carried out the preliminary work. SSH analyzed the results and wrote the primary draft of the article. All authors critically reviewed, edited, and approved the final manuscript.

Conflict of Interests

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

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, The Lancet infectious diseases (2020).

[2] M. Boni, P. Lemey, X. Jiang, T. Lam, B. Perry, T. Castoe, A. Rambaut, D. Robertson, Evolutionary origins of the sars-cov-2 sarbecovirus lineage responsible for the covid-19 pandemic. nat microbiol (2020).

[3] K. Ohnuki, M. Yoshimoto, H. Fujii, Radiological protection and biological covid-19 protection in the nuclear medicine department, European Journal of Nuclear Medicine and Molecular Imaging (2020) 1–3.

[4] A. E. R. M. P. S. E. L. Dediego ML, Pewe L, Pathogenicity of severe acute respiratory coronavirus deletion mutants in hace-2 transgenic mice, Virology (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, Cellular and Molecular Life Sciences (2020) 1–34.

[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, et al., Covid-19 disease map, building a computational repository of sars-cov-2 virus-host interaction mechanisms, Scientific data 7 (1) (2020) 1–4.

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

[8] S. Ozono, Y. Zhang, H. Ode, T. T. Seng, K. Imai, K. Miyoshi, S. Kishigami, T. Ueno, Y. Iwatani, T. Suzuki, et al., Naturally mutated spike proteins of sars-cov-2 variants show differential levels of cell entry, bioRxiv (2020).

[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 (2020).

[10] M. R. Islam, M. N. Hoque, M. S. Rahman, A. R. U. 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, Scientific reports 10 (1) (2020) 1–9.

[11] Y. Ren, T. Shu, D. Wu, J. Mu, C. Wang, M. Huang, Y. Han, X.-Y. Zhang, W. Zhou, Y. Qiu, et al., The orf3a protein of sars-cov-2 induces apoptosis in cells, Cellular & molecular immunology 17 (8) (2020) 881–883.

[12] 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 (2020).

[13] 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).

[14] A. Shah, Novel coronavirus-induced nlrp3 inflammasome activation: A potential drug target in the treatment of covid-19, Frontiers in Immunology 11 (2020).

[15] 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 research 286 (2020) 198074.

[16] 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 research notes 4 (1) (2011) 1–9.

[17] H. Xia, Z. Cao, X. Xie, X. Zhang, J. Y.-C. Chen, H. Wang, V. D. Menachery, R. Rajsbaum, P.-Y. Shi, Evasion of type i interferon by sars-cov-2, Cell Reports 33 (1) (2020) 108234.

[18] 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 (jan-mar 2020), Journal of virology (2020).

[19] X. Lei, X. Dong, R. Ma, W. Wang, X. Xiao, Z. Tian, C. Wang, Y. Wang, L. Li, L. Ren, et al., Activation and evasion of type i interferon responses by sars-cov-2, Nature communications 11 (1) (2020) 1–12.

[20] F. Pereira, Evolutionary dynamics of the sars-cov-2 orf8 accessory gene, Infection, Genetics and Evolution 85 (2020) 104525.

[21] S. S. Hassan, S. Ghosh, D. Attrish, P. P. Choudhury, M. Seyran, D. Pizzol, P. Adadi, T. M. Abd El Aziz, A. Soares, R. Kandimalla, et al., A unique view of sars-cov-2 through the lens of orf8 protein, bioRxiv (2020).

[22] Y. C. Su, D. E. Anderson, B. E. Young, M. Linster, F. Zhu, J. Jayakumar, Y. Zhuang, S. Kalimuddin, J. G. Low, C. W. Tan, et al., 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).

[23] R. Cagliani, D. Forni, M. Clerici, M. Sironi, Coding potential and sequence conservation of sars-cov-2 and related animal viruses, Infection, Genetics and Evolution (2020) 104353.

[24] S. S. Hassan, D. Attrish, S. Ghosh, P. P. Choudhury, V. N. Uversky, B. D. Uhal, K. Lundstrom, N. Rezaei, A. A. Aljabali, M. Seyran, et al., Notable sequence homology of the orf10 protein introspects the architecture of sars-cov-2, bioRxiv (2020).

[25] F. K. Yoshimoto, The proteins of severe acute respiratory syndrome coronavirus-2 (sars cov-2 or n-cov19), the cause of covid-19, The Protein Journal (2020) 1.

[26] K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker, Z. Obradovic, Optimizing long intrinsic disorder predictors with protein evolutionary information, Journal of bioinformatics and computational biology 3 (01) (2005) 35–60.

[27] Z.-L. Peng, L. Kurgan, Comprehensive comparative assessment of in-silico predictors of disordered regions, Current Protein and Peptide Science 13 (1) (2012) 6–18.

[28] X. Fan, L. Kurgan, Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus, Journal of Biomolecular Structure and Dynamics 32 (3) (2014) 448–464.

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

[30] 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 and Research Perspectives 11 (3) (2020) 101.

[31] M. Seyran, D. Pizzol, P. Adadi, T. M. A. El-Aziz, S. S. Hassan, A. Soares, R. Kandimalla, K. Lundstrom, M. Tambuwala, A. A. Aljabali, et al., Questions concerning the proximal origin of sars-cov-2, Journal of Medical Virology (2020).

[32] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for sars-cov-2, Nature Reviews Microbiology (2020) 1–16.

[33] L. M. Brufsky A, Ratcheting down the virulence of sars-cov-2 in the covid-19 pandemic, Journal of Medical Virology (2020).

Supplementary Data

Table 9: List of ORF6 sequences and their accessions

Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession
S1	QNE11892	S13	QKO25582	S25	QOH29849	S37	QKG86850	S49	QLB39372	S61	QNO87603
S2	QOI60338	S14	QJS54110	S26	QNA37799	S38	QNA37703	S50	QOF13773	S62	QKV37732
S3	QMX85113	S15	QJC19423	S27	QNR99459	S39	QJA16812	S51	QOH27322	S63	QNO58695
S4	QMU94792	S16	QOI53465	S28	QOE87928	S40	QNM80965	S52	QKQ63440	S64	QNO87963
S5	QIU81889	S17	QJT72174	S29	QMT93272	S41	QOC65745	S53	QKV39531	S65	QNO67179
S6	QNJ45330	S18	QJC21021	S30	QOC65685	S42	QKU31089	S54	QKU30333	S66	QLG76377
S7	QNL36014	S19	QJT72858	S31	QMJ19995	S43	QNU10840	S55	QNN87740	S67	QKV37240
S8	QMU94900	S20	QMI98359	S32	QLI50453	S44	QOF10845	S56	QMS54339	S68	QNO62835
S9	YP_009724394	S21	QOI10363	S33	QOF08397	S45	QOF14025	S57	QMJ00949	S69	QJR87301
S10	QNL98449	S22	QMJ00613	S34	QKV39052	S46	QLC91284	S58	QMT49529	S70	QNV49474
S11	QKO25642	S23	QMU25387	S35	QOJ86810	S47	QOF12309	S59	QNP00779	S71	QNV50338
S12	QKV49390	S24	QNL13170	S36	QMT57060	S48	QKV06503	S60	QJR87841	S72	QMB22615

Table 10: List of ORF7a sequences and their accessions

Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession
S1	QKT21007	S27	QJF76096	S53	QNL23941	S79	QKU33322	S105	QOF20590	S131	QNO59248
S2	QMX85114	S28	QOF09958	S54	QOJ40868	S80	QMU25172	S106	QNS28602	S132	QNO66520
S3	QNN90081	S29	QNQ16882	S55	QOF08182	S81	QMS51487	S107	QOF19618	S133	QNO90124
S4	QOI60339	S30	QLH27803	S56	QNA38004	S82	QMU25112	S108	QMS54508	S134	QNO93328
S5	QJZ28119	S31	QOH26975	S57	QOF14542	S83	QNL11851	S109	QIZ14130	S135	QNO64828
S6	QMX85090	S32	QNO32893	S58	QNC69326	S84	QMJ00890	S110	QNM94399	S136	QNO73996
S7	QMU94793	S33	QOF21250	S59	QNL23738	S85	QLG00044	S111	QNS17749	S137	QKV37313
S8	QLH56163	S34	QLJ93776	S60	QOE87249	S86	QMT53402	S112	QLG98244	S138	QNO88516
S9	QLH56283	S35	QOI11612	S61	QOC60864	S87	QOI10196	S113	QNS17689	S139	QJR92882
S10	QLR07201	S36	QMI94146	S62	QMT93717	S88	QNU10973	S114	QNI23442	S140	QNO58696
S11	QLH90095	S37	QJR84962	S63	QNS29326	S89	QNA39599	S115	QMT48978	S141	QNO92404
S12	QNB17764	S38	QMI91915	S64	QOF21190	S90	QOE87321	S116	QMT48966	S142	QLG76138
S13	QNL98462	S39	QNS30322	S65	QNG41678	S91	QNN95802	S117	QOC61296	S143	QNO90292
S14	QJW69144	S40	QNI24809	S66	QNS28494	S92	QMT91425	S118	QMJ01226	S144	QNO58720
S15	QNL35871	S41	QOI60423	S67	QNN95250	S93	QMT91245	S119	QMX86011	S145	QNO64900
S16	YP_009724395	S42	QNT35437	S68	QMU94037	S94	QMU94709	S120	QNM94471	S146	QLG75826
S17	QMS95046	S43	QNI23370	S69	QOE87309	S95	QOE76015	S121	QKS90016	S147	QNO61324
S18	QLF97752	S44	QLC46450	S70	QOJ86811	S96	QOE81298	S122	QIX13980	S148	QNO70240
S19	QKQ30155	S45	QOE44625	S71	QNU10477	S97	QKV41188	S123	QOE81166	S149	QNO80368
S20	QNJ45535	S46	QKU53354	S72	QNQ16954	S98	QOH29130	S124	QLJ58162	S150	QNO66460
S21	QLA46617	S47	QMI90741	S73	QIS61307	S99	QJX74540	S125	QOI60459	S151	QNO96076
S22	QNR60421	S48	QKU33310	S74	QJS54699	S100	QJF76899	S126	QIU81254	S152	QNP02976
S23	QKY65318	S49	QOF07894	S75	QND76351	S101	QMT51066	S127	QNO69016	S153	QNO59080
S24	QJR84386	S50	QMI97185	S76	QOF12514	S102	QNS30406	S128	QNP05940	S154	QNV50135
S25	QMT97821	S51	QNV71003	S77	QNQ17002	S103	QLC48045	S129	QNO69448	S155	QNV50219
S26	QOI53466	S52	QOI10028	S78	QNN86414	S104	QOF08242	S130	QNP07212	S156	QNV49775
										S157	QNV49475
										S158	QNV49691

Table 7: List of ORF3a Sequences and their corresponding accessions

Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession
S1	QKS66941	S32	QMT97817	S63	QJQ04482	S94	QJS54155	S125	QNN96349	S156	QNN87761	S187	QOE87817	S218	QOH29606
S2	QOI60359	S33	QOI53581	S64	QLH56279	S95	QJT72471	S126	QNU11125	S157	QNI24517	S188	QOF15546	S219	QMI95294
S3	QKR84274	S34	QJR84430	S65	QLF97736	S96	QJT72387	S127	QOF11478	S158	QJD23478	S189	QKG88539	S220	QOE87662
S4	QLQ87565	S35	QJX44407	S66	QMU94885	S97	QNT10049	S128	QNS40120	S159	QOF08730	S190	QNN87845	S221	QMS51926
S5	QOI60335	S36	QNN90029	S67	QKO25735	S98	QNN93028	S129	QMU92084	S160	QNE13005	S191	QJD47551	S222	QNN95210
S6	QMX85002	S37	QLH93453	S68	QLA10093	S99	QIZ16548	S130	QOF14346	S161	QKE45885	S192	QJD25758	S223	QKG91107
S7	QMX85026	S38	QLH56231	S69	QKK12852	S100	QJS53687	S131	QNO40166	S162	QKS65621	S193	QMI94418	S224	QMU25960
S8	QNL98458	S39	QJW00412	S70	QKY59990	S101	QJS54215	S132	QLJ93700	S163	QNG41554	S194	QNO31749	S225	QOH26275
S9	QKS67456	S40	QLQ87577	S71	QLF98084	S102	QKE10935	S133	QOC62516	S164	QLM05764	S195	QNU10705	S226	QOC67026
S10	QLH93429	S41	QKO25747	S72	QLF98048	S103	QJS39497	S134	QMT49694	S165	QNU10765	S196	QNA36464	S227	QNU07492
S11	QLF98036	S42	QKX47995	S73	QLH56099	S104	QNC69819	S135	QKV40164	S166	QNS00146	S197	QLY88564	S228	QIZ16438
S12	QNL35975	S43	QMS50988	S74	QMU94765	S105	QOE86813	S136	QOH26347	S167	QMI95795	S198	QKG90399	S229	QOH27307
S13	QKE61733	S44	QKU37034	S75	QNL36011	S106	QJS53735	S137	QNU11783	S168	QKG64052	S199	QOF18738	S230	QKG86518
S14	QNL35963	S45	QLL26047	S76	QJY40506	S107	QLJ53549	S138	QNL11079	S169	QJW28665	S200	QOF11886	S231	QOH26287
S15	QLH93441	S46	QKJ84956	S77	QMJ01246	S108	QJS54191	S139	QOH26407	S170	QLA47500	S201	QNN96433	S232	QOI09856
S16	QNT09953	S47	QHZ00380	S78	QLH93202	S109	QJW69023	S140	QNL12447	S171	QKN20812	S202	QOH26875	S233	QNL11403
S17	QLF97772	S48	QJD47873	S79	QLH55768	S110	QJS39520	S141	QJV21807	S172	QMU25384	S203	QOE75507	S234	QMT27641
S18	QLF97952	S49	QJD47897	S80	QLQ87733	S111	QNO10703	S142	QMI91059	S173	QMT28157	S204	QNA38024	S235	QOF14070
S19	QLR12406	S50	QNB17760	S81	QLQ87613	S112	QNJ45107	S143	QKG81932	S174	QNI24121	S205	QOC65862	S236	QNO40082
S20	QJX44383	S51	QLF98261	S82	QLA10165	S113	QLF78310	S144	QOH26143	S175	QOC66030	S206	QOD06326	S237	QOC65886
S21	QNL90873	S52	QMU94777	S83	QLH55720	S114	QOI53509	S145	QOF10062	S176	QOC62960	S207	QNS27437	S238	QKU28847
S22	QJD47849	S53	QKI31239	S84	QNN88251	S115	QJT72951	S146	QMT92621	S177	QOD06902	S208	QKV06224	S239	QMI93734
S23	QJW00292	S54	QKK14624	S85	QNN88071	S116	QKS66053	S147	QOF16914	S178	QLI50222	S209	QNA36908	S240	QNO40418
S24	QLF98201	S55	QLH55816	S86	QJW69308	S117	QJD47419	S148	QNV70291	S179	QOH26179	S210	QOD07286	S241	QOJ86807
S25	QMS51324	S56	QLF97844	S87	QNT10169	S118	QJC19648	S149	QKV41616	S180	QMU26188	S211	QOC60968	S242	QMU91460
S26	QMJ01294	S57	QMU94981	S88	QLC48564	S119	QNM94191	S150	QNN87893	S181	QOJ41128	S212	QOH29786	S243	QOI60419
S27	QKI28662	S58	QMU84947	S89	QNT10181	S120	QNR51233	S151	QOI60263	S182	QNA36548	S213	QNA39116	S244	QNA37028
S28	QLH56255	S59	QLH55840	S90	QJT72507	S121	QNS28466	S152	QMI96608	S183	QOE87233	S214	QJI07211	S245	QNS00978
S29	QNN88131	S60	YP_009724391	S91	QJT72363	S122	QOF12666	S153	QNI23378	S184	QNA40027	S215	QKV26659	S246	QOC63200
S30	QLG75126	S61	QJD20838	S92	QJT72327	S123	QMT55653	S154	QLP89461	S185	QIZ13838	S216	QOH26791	S247	QKG90495
S31	QLQ87661	S62	QMJ01306	S93	QOI53462	S124	QNS27905	S155	QMU94453	S186	QNN87929	S217	QOF16302	S248	QOD06554

Table 8: Contd . . . List of ORF3a Sequences and their corresponding accessions

Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession
S249	QOI10024	S291	QNI25093	S333	QOC62312	S375	QNM94071	S417	QOE75567	S459	QNO91800	S501	QLF80217
S250	QOI09700	S292	QNE12105	S334	QOC61580	S376	QJE38451	S418	QNU12487	S460	QKV37633	S502	QNV49951
S251	QOF08166	S293	QNA36128	S335	QOI60491	S377	QKS67001	S419	QKN19672	S461	QKV38005	S503	QNV50107
S252	QOF13494	S294	QOF09318	S336	QOE45156	S378	QOC66078	S420	QNN86050	S462	QNO78132	S504	QNV50443
S253	QOI09832	S295	QOI09952	S337	QMT54658	S379	QMI94238	S421	QLL35972	S463	QKV38209	S505	QNV49471
S254	QOE75987	S296	QMJ00874	S338	QKX46204	S380	QLG98012	S422	QOC61460	S464	QLG75930	S506	QNV49711
S255	QOJ87728	S297	QMJ01126	S339	QJR84790	S381	QMT54142	S423	QMT91241	S465	QKV38257	S507	QNV49723
S256	QNU12141	S298	QMU94705	S340	QLH01382	S382	QJI53932	S424	QOH27247	S466	QNO61164	S508	QNV49651
S257	QOJ42000	S299	QNO30433	S341	QMS53076	S383	QMT53842	S425	QOI09964	S467	QNO72312	S509	QNH88920
S258	QOF16962	S300	QMS53136	S342	QLJ57252	S384	QKE44990	S426	QMT91985	S468	QNP01544	S510	QNV50467
S259	QMJ00754	S301	QMU93169	S343	QNS27305	S385	QNI24649	S427	QNR99456	S469	QNO62280		
S260	QNA38792	S302	QMU93793	S344	QJU11554	S386	QNU10485	S428	QMT48962	S470	QNO98808		
S261	QJS57052	S303	QKE45861	S345	QOF13194	S387	QOC66942	S429	QNN96086	S471	QNO95016		
S262	QNS17793	S304	QMI99855	S346	QMJ00970	S388	QOI10216	S430	QNL11703	S472	QNO64260		
S263	QOI09604	S305	QMS52034	S347	QKU31182	S389	QOC61880	S431	QOI11812	S473	QJR88078		
S264	QNA39104	S306	QMT49538	S348	QJX70592	S390	QNV70267	S432	QMU91328	S474	QNO74772		
S265	QND77223	S307	QOF17214	S349	QOF11826	S391	QOH29042	S433	QKW88844	S475	QNO97920		
S266	QLG97532	S308	QII57239	S350	QLC92421	S392	QJD23730	S434	QNO30469	S476	QNO67524		
S267	QIS61075	S309	QKU53854	S351	QMI97840	S393	QOF13590	S435	QIU81286	S477	QNO69120		
S268	QJW28449	S310	QNG41770	S352	QKC05357	S394	QLL36314	S436	QLK98276	S478	QNO94488		
S269	QLC91905	S311	QNG41494	S353	QOJ39681	S395	QNM80986	S437	QKV07400	S479	QNO61584		
S270	QNA37196	S312	QMI98428	S354	QNG41458	S396	QMU91484	S438	QLY90498	S480	QNP05276		
S271	QIZ15958	S313	QNL11643	S355	QLJ57600	S397	QMT57105	S439	QNA36272	S481	QNP05804		
S272	QOF21390	S314	QOF11550	S356	QOJ51963	S398	QNO30373	S440	QOC65706	S482	QNO58716		
S273	QNI24985	S315	QNS30222	S357	QOC60800	S399	QNN87941	S441	QOJ40996	S483	QNO62304		
S274	QMI95078	S316	QMJ00382	S358	QMT96581	S400	QMT49358	S442	QMI93663	S484	QNO58692		
S275	QNS27233	S317	QMS54228	S359	QKV06236	S401	QOC61400	S443	QJR88390	S485	QNP06092		
S276	QOC66378	S318	QKV07340	S360	QNN86122	S402	QNJ45603	S444	QJR88102	S486	QNP06020		
S277	QOI10108	S319	QMI94610	S361	QLJ58038	S403	QKU32982	S445	QJR88822	S487	QLG76386		
S278	QOH27055	S320	QOC61328	S362	QOF17190	S404	QJA17681	S446	QNO61488	S488	QNO83328		
S279	QOI11752	S321	QKU29051	S363	QKG90867	S405	QLC94305	S447	QKV38281	S489	QNO77160		
S280	QKN20824	S322	QOI11644	S364	QNA38000	S406	QNR99516	S448	QNO61728	S490	QNO62496		
S281	QOF14322	S323	QOF09342	S365	QOC60872	S407	QNI24613	S449	QNO81756	S491	QNO61344		
S282	QKV41592	S324	QOC61832	S366	QOC62540	S408	QOJ41728	S450	QJR88306	S492	QLG75678		
S283	QOF19914	S325	QNM94431	S367	QLI51782	S409	QLJ58146	S451	QJR89110	S493	QNP03152		
S284	QOF09366	S326	QMI93255	S368	QOF11250	S410	QOE87557	S452	QNO76092	S494	QNP01832		
S285	QNM81142	S327	QOC61352	S369	QNN95714	S411	QNM94131	S453	QLG76542	S495	QLG75822		
S286	QNN95246	S328	QOF20406	S370	QIS61315	S412	QOH29630	S454	QNO63024	S496	QNV49855		
S287	QNL12159	S329	QLC92601	S371	QMT55089	S413	QLH57751	S455	QNO91668	S497	QNV49531		
S288	QJD47203	S330	QKK14612	S372	QJF77147	S414	QNE12273	S456	QJR95110	S498	QNV50491		
S289	QNS27329	S331	QKU28463	S373	QLK98372	S415	QKG81824	S457	QNO93300	S499	QNV50119		
S290	QMJ01054	S332	QOI60287	S374	QNL24069	S416	QOJ51999	S458	QLG75942	S500	QNV49999		

Table 11: List of ORF7b sequences and their accessions

Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession
S1	QNN90094	S11	QOI53467	S21	QOJ86812	S31	QNP01153
S2	QOI60340	S12	QNU10478	S22	QMI93943	S32	QLG75935
S3	QKX48960	S13	QNL24002	S23	QOH26412	S33	QNO58697
S4	QNJ45416	S14	QNL11060	S24	QNV70236	S34	QNO90425
S5	QMU84916	S15	QNU10274	S25	QNO30378	S35	QLG75923
S6	YP_009725318	S16	QMT53511	S26	QKE45866	S36	QNO74885
S7	QNN88304	S17	QOI10365	S27	QJC19833	S37	QNV49476
S8	QJQ84777	S18	QOF08027	S28	QKU28444		
S9	QNL90926	S19	QMI96613	S29	QJD47604		
S10	QKM76816	S20	QKY77886	S30	QKV38827		

Table 13: List of ORF10 Sequences and their corresponding accessions

Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession
S1	QOI60343	S12	QKM76363	S23	QOF10934	S34	QNQ16982
S2	QIS29991	S13	QNI23218	S24	QNO31049	S35	QOF07898
S3	QNR60413	S14	QOJ41112	S25	QOH29638	S36	QOD06766
S4	QNC49349	S15	QNC04532	S26	QOF11654	S37	QNO92660
S5	QNJ45359	S16	QLA48060	S27	QOJ86815	S38	QKV37377
S6	QNN88665	S17	QOE87730	S28	QKV08176	S39	QKV37245
S7	YP_009725255	S18	QOJ41016	S29	QOI09960	S40	QNO58700
S8	QNB17780	S19	QOF17054	S30	QOF10718	S41	QNO92552
S9	QKJ68385	S20	QOE87313	S31	QLG99793	S42	QNO73604
S10	QOE86821	S21	QKU54102	S32	QOH29566	S43	QNV49479
S11	QOI53470	S22	QNI25281	S33	QOE87969	S44	QNV50343

Table 12: List of ORF8 Sequences and their corresponding accessions

Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession	Seq Name	Accession
S1	QOI60353	S26	QKJ68684	S51	QMU91718	S76	QNQ15783	S101	QKU37544	S126	QOE45246	S151	QNA36434	S176	QJR88180
S2	QKR83992	S27	QNB42358	S52	QMT50804	S77	QMT55731	S102	QMI92229	S127	QNE12363	S152	QLH57924	S177	QNO58698
S3	QOI60341	S28	QLF97742	S53	QNA37310	S78	QOE87943	S103	QMT55779	S128	QOI11698	S153	QKV39247	S178	QNP06014
S4	QNN90083	S29	QNN88233	S54	QOH29180	S79	QOH27313	S104	QOH26893	S129	QLH01196	S154	QNA41569	S179	QNP06302
S5	QMX85116	S30	QLQ87643	S55	QNN87191	S80	QMT49652	S105	QJC19630	S130	QOI09922	S155	QOF08352	S180	QNO92466
S6	QJY40584	S31	QNR60016	S56	QNT54575	S81	QLK98030	S106	QNT08732	S131	QOC66060	S156	QOF07872	S181	QJR88996
S7	QMU94963	S32	QNT10199	S57	QJS56890	S82	QNL36557	S107	QMJ01228	S132	QNL11313	S157	QNU10627	S182	QNO80454
S8	QLH64870	S33	QLF78328	S58	QNO32391	S83	QNQ17016	S108	QOD07160	S133	QMU91334	S158	QOH27949	S183	QJR93136
S9	QMU94975	S34	QJT72369	S59	QNV70369	S84	QOF07896	S109	QOJ87257	S134	QMT50180	S159	QNL24003	S184	QNO78834
S10	QKU37052	S35	QOI53468	S60	QOE80988	S85	QNN95156	S110	QOI60413	S135	QNU12589	S160	QMU92030	S185	QNO64554
S11	QJX44617	S36	QJS53933	S61	QOJ51993	S86	QJS57274	S111	QNA40801	S136	QNN86440	S161	QNN95456	S186	QJR93160
S12	QKV49393	S37	QNO10661	S62	QOJ86813	S87	QJC19618	S112	QMJ01216	S137	QMU94207	S162	QNM94077	S187	QNV50269
S13	QMS95048	S38	QJS53513	S63	QNN96092	S88	QNS30312	S113	QNU10771	S138	QNL36437	S163	QNO31203	S188	QNV49477
S14	QKQ29929	S39	QNC69777	S64	QOC62438	S89	QNL11769	S114	QOE78974	S139	QJX70307	S164	QOC61814	S189	QNV49633
S15	QOI53587	S40	QJZ28306	S65	QLH58953	S90	QKV06506	S115	QJW28311	S140	QNU12087	S165	QOC60986	S190	QNV50437
S16	QLF97850	S41	QLJ84632	S66	QNA41641	S91	QNO39956	S116	QNS01609	S141	QNO31095	S166	QNO87282		
S17	QMU84881	S42	QNL11649	S67	QNA37358	S92	QNS26661	S117	QOH27661	S142	QNL11961	S167	QNP02198		
S18	QMU84893	S43	QMT48896	S68	QMJ00652	S93	QKV40062	S118	QNA41713	S143	QOH29024	S168	QNO61410		
S19	YP_009724396	S44	QOF13404	S69	QNL24063	S94	QLA47566	S119	QNN95252	S144	QMT56763	S169	QNO91674		
S20	QJT43615	S45	QMT28672	S70	QLC47867	S95	QMI96758	S120	QOJ51981	S145	QMT50144	S170	QNO92022		
S21	QNL35933	S46	QMS51860	S71	QNO31407	S96	QNO30439	S121	QND77013	S146	QNO31371	S171	QNO75774		
S22	QKI36860	S47	QOF09360	S72	QLM05830	S97	QMI91809	S122	QJD48694	S147	QNL23800	S172	QKV38119		
S23	QNN88663	S48	QNS00212	S73	QOI10210	S98	QOE87299	S123	QMS54342	S148	QNO40328	S173	QKV37315		
S24	QNL98464	S49	QMS53022	S74	QLJ93922	S99	QOD06728	S124	QMT96539	S149	QLB39040	S174	QNO98598		
S25	QLE11192	S50	QNR51095	S75	QNA38594	S100	QMT95795	S125	QOI10114	S150	QOE45090	S175	QNO92610		

Color Schemes																		
			Africa			Asia		Europe	N. America		Ocenia		S. America					
ORF3a																		
S1	S38	S85	S121	S146	S171	S192	S217	S238	S260	S282	S306	S336	S361	S389	S416	S441	S476	
S3	S39	S86	S123	S147	S172	S193	S218	S239	S261	S283	S307	S337	S362	S392	S417	S442	S477	
S6	S40	S87	S125	S149	S173	S194	S219	S240	S262	S284	S308	S339	S363	S393	S418	S443	S481	
S10	S41	S90	S126	S150	S174	S195	S220	S242	S263	S285	S310	S340	S365	S394	S419	S445	S483	
S11	S42	S96	S128	S151	S175	S197	S221	S243	S264	S286	S311	S342	S367	S397	S420	S449	S485	
S14	S43	S97	S129	S152	S176	S198	S222	S244	S265	S287	S313	S343	S368	S398	S421	S451	S487	
S15	S44	S98	S130	S153	S177	S200	S223	S245	S267	S288	S316	S345	S370	S399	S422	S453	S488	
S16	S45	S99	S131	S154	S178	S201	S224	S246	S268	S290	S317	S346	S371	S400	S423	S457	S489	
S18	S47	S102	S132	S156	S179	S202	S225	S247	S269	S291	S318	S347	S372	S402	S425	S458	S490	
S19	S55	S106	S133	S157	S180	S203	S226	S248	S270	S292	S320	S348	S373	S403	S426	S460	S494	
S20	S56	S108	S134	S158	S181	S204	S228	S249	S272	S293	S324	S349	S374	S404	S427	S461	S495	
S21	S57	S109	S135	S159	S182	S205	S229	S250	S273	S296	S326	S350	S375	S405	S428	S462	S498	
S22	S61	S110	S136	S160	S183	S206	S230	S251	S274	S297	S327	S351	S376	S406	S429	S463	S499	
S24	S64	S112	S138	S161	S184	S207	S231	S252	S275	S298	S328	S352	S377	S407	S430	S464	S506	
S29	S73	S114	S139	S162	S185	S209	S232	S253	S276	S299	S329	S354	S379	S409	S433	S465	S510	
S30	S75	S116	S140	S164	S187	S211	S233	S254	S277	S300	S330	S355	S380	S411	S435	S467		
S31	S76	S117	S141	S165	S188	S213	S234	S255	S278	S301	S331	S357	S381	S412	S436	S469		
S32	S78	S118	S142	S168	S189	S214	S235	S256	S279	S302	S332	S358	S384	S413	S437	S470		
S35	S80	S119	S143	S169	S190	S215	S236	S257	S280	S303	S333	S359	S386	S414	S438	S471		
S36	S81	S120	S144	S170	S191	S216	S237	S259	S281	S305	S335	S360	S387	S415	S439	S475		
ORF6			ORF7a					ORF7b	ORF8						ORF10			
S1	S31	S59	S3	S39	S72	S96	S119	S4	S5	S46	S75	S99	S124	S152	S2	S34		
S3	S32	S61	S5	S40	S73	S97	S121	S7	S6	S47	S76	S102	S125	S153	S3	S35		
S4	S33	S62	S6	S42	S74	S98	S122	S9	S10	S49	S77	S103	S126	S154	S5	S36		
S5	S34	S68	S8	S43	S76	S99	S123	S12	S13	S52	S78	S104	S127	S155	S8	S39		
S6	S37	S69	S9	S44	S77	S100	S125	S14	S18	S53	S79	S105	S129	S157	S13	S41		
S8	S40		S11	S45	S78	S101	S126	S15	S20	S55	S80	S106	S131	S158	S14	S42		
S11	S41		S17	S50	S79	S102	S127	S16	S22	S56	S81	S107	S132	S159	S15	S44		
S13	S42		S19	S51	S80	S103	S128	S17	S23	S57	S82	S110	S133	S160	S16			
S14	S44		S20	S53	S82	S104	S132	S20	S27	S58	S85	S111	S135	S161	S17			
S15	S45		S22	S55	S83	S105	S133	S22	S28	S60	S86	S112	S136	S162	S19			
S18	S46		S23	S56	S84	S106	S136	S23	S30	S61	S88	S113	S137	S163	S20			
S20	S47		S27	S58	S85	S108	S143	S25	S31	S63	S90	S114	S139	S164	S21			
S21	S48		S28	S59	S86	S109	S148	S26	S33	S64	S91	S115	S141	S165	S22			
S22	S49		S29	S60	S87	S110	S150	S27	S34	S65	S92	S116	S142	S167	S24			
S24	S51		S30	S62	S88	S112	S153	S28	S36	S66	S93	S117	S144	S175	S25			
S25	S53		S31	S63	S89	S114	S154	S29	S41	S67	S94	S118	S145	S179	S28			
S26	S54		S32	S65	S91	S115		S30	S42	S69	S95	S119	S146	S180	S30			
S28	S55		S33	S66	S92	S116		S32	S43	S70	S96	S121	S148	S184	S31			
S29	S56		S36	S68	S94	S117		S35	S44	S73	S97	S122	S150	S186	S32			
S30	S57		S37	S71	S95	S118		S36	S45	S74	S98	S123	S151	S187	S33			

Figure 11: Unique variants of accessory protein sequences.

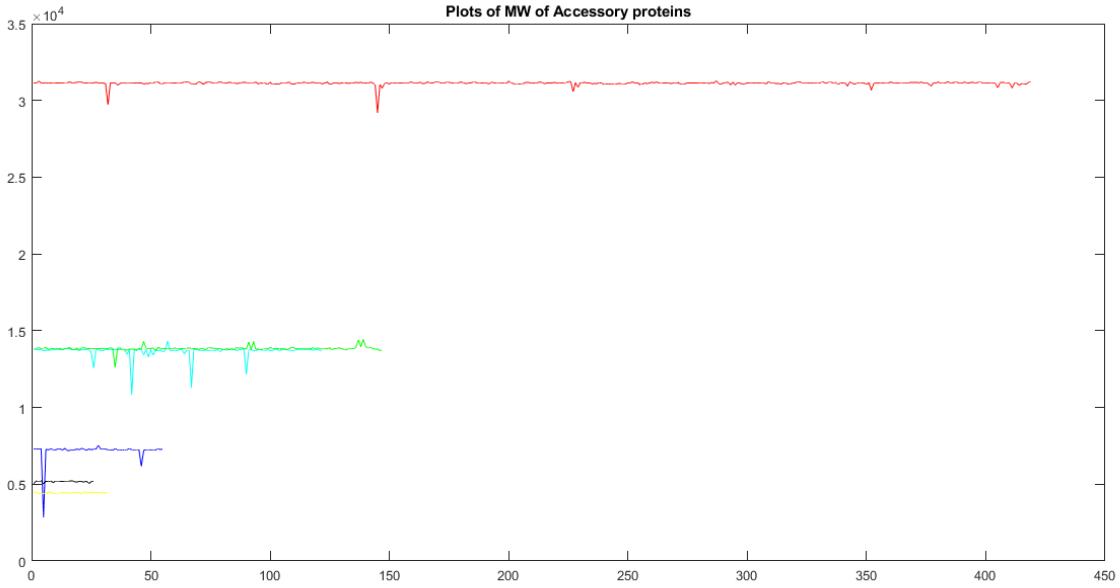


Figure 12: Graphical representations of molecular weights of unique accessory proteins

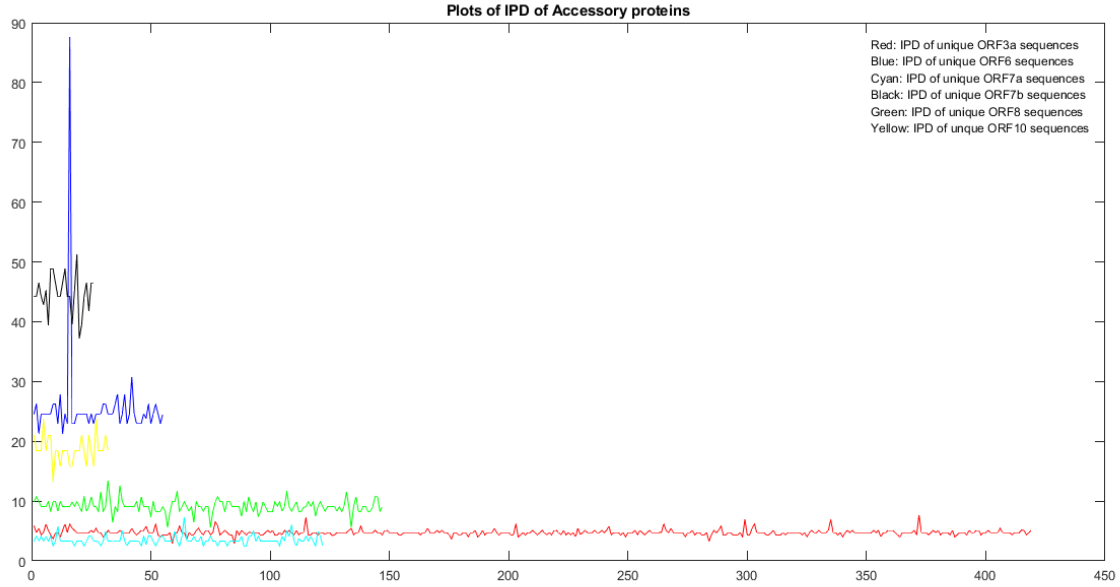


Figure 13: Graphical representations of Intrinsic Disorder contents of unique accessory proteins

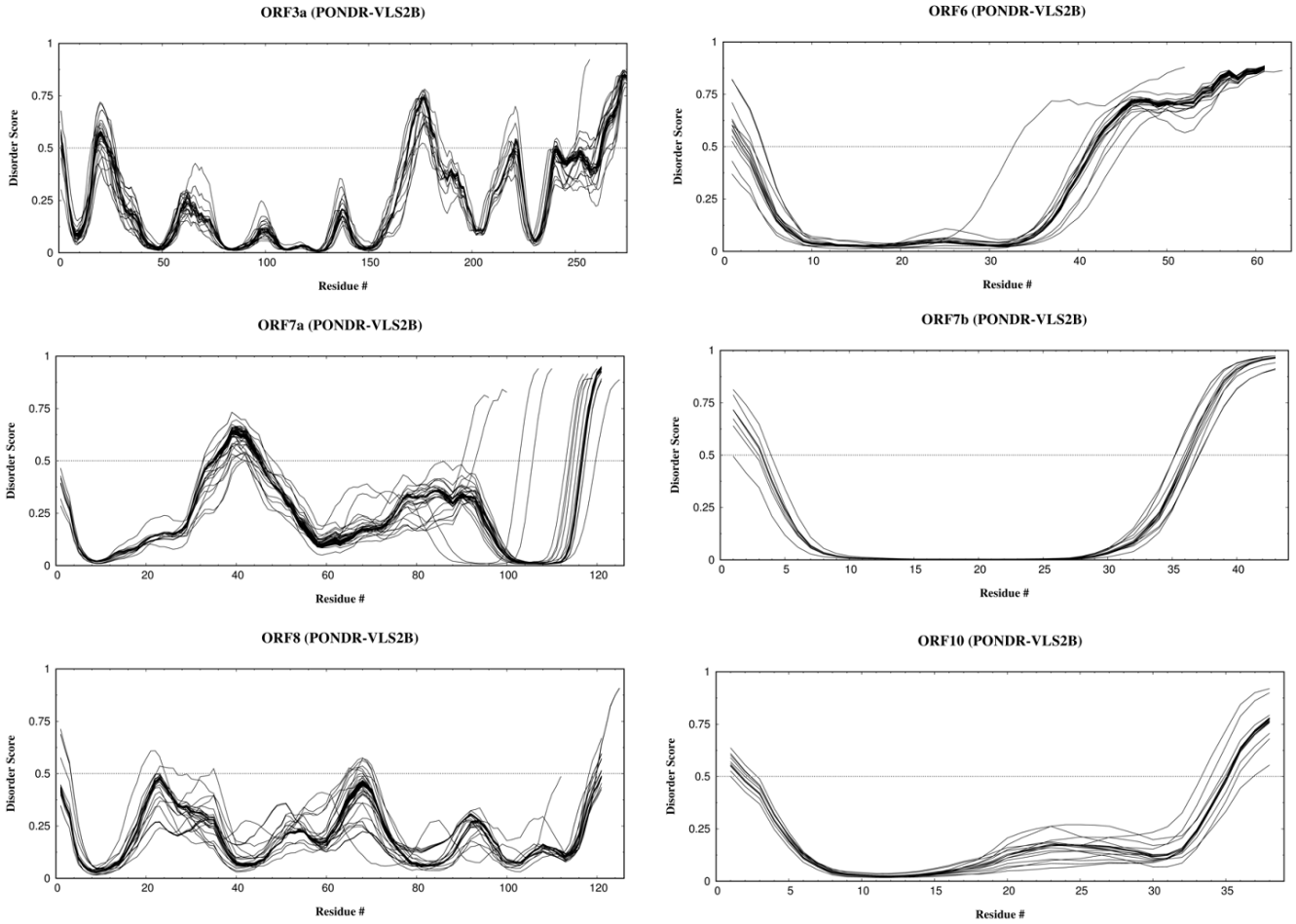


Figure 14: disorder plots of all unique variants for six accessory proteins of SARS-CoV-2

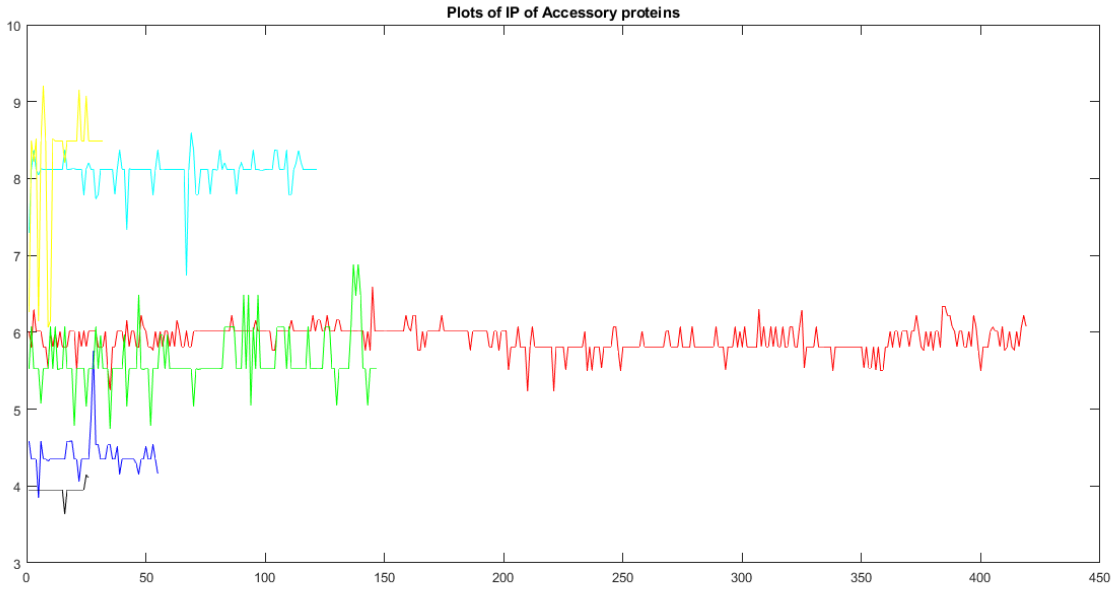


Figure 15: Graphical representations of isoelectric points of unique accessory proteins