

Type of the Paper (Review)

SARS-CoV-2 ORF8 and SARS-CoV ORF8ab: Genomic Divergence and Functional Convergence

Sameer Mohammad¹, Abderrezak Bouchama¹, Bothina Mohammad Alharbi¹, Mamoon Rashid², Tanveer Saleem Khatlani³, Nusaibah S Gaber¹, and Shuja Shafi Malik^{1,*}

¹King Abdullah International Medical Research Center/King Saud bin Abdulaziz University for Health Sciences, Experimental Medicine Department-MNGHA, Riyadh, 11426, Saudi Arabia

²King Abdullah International Medical Research Center/King Saud bin Abdulaziz University for Health Sciences, Bioinformatics and Biostatistics Department-MNGHA, Riyadh, 11426, Saudi Arabia

³King Abdullah International Medical Research Center/King Saud bin Abdulaziz University for Health Sciences, Stem Cells Unit, Department of Cellular Therapy-MNGHA, Riyadh, 11426, Saudi Arabia

mohammadsa1@ngha.med.sa(S.M), bouchamaab@ngha.med.sa(A.B), bothinaalharbi@hotmail.com(B. M.A), rashidma@ngha.med.sa (M.R), khatlanita@ngha.med.sa(T.S.K), nusaibahgaber@gmail.com (N.S.G)

* Correspondence: maliksh@ngha.med.sa (S.S.M)

Abstract: COVID-19 pandemic in first seven months has led to more than 15 million confirmed infected cases and 600,000 deaths. SARS-CoV-2, the causative agent for COVID-19 has proved a great challenge for its ability to spread in asymptomatic stages and a diverse disease spectrum it has generated. This has created a challenge of unimaginable magnitude not only affecting human health and life but also potentially generating a long-lasting socioeconomic impact. Both medical sciences and biomedical research have also been challenged consequently leading to a large number of clinical trials and vaccine initiatives. While known proteins of pathobiological importance are targets for these therapeutic approaches, it is imperative to explore other factors of viral significance. Accessory proteins are one such trait that have diverse roles in coronavirus pathobiology. Here we analyze certain genomic characteristics of SARS-CoV-2 accessory protein ORF8, predict upon its protein features and review current available literature regarding its function. We have also undertaken review of ORF8 homolog ORF8ab from SARS-CoV with a purpose of developing holistic understanding of these proteins for reason that coronaviruses have been infecting humans repeatedly and might continue to do so. Despite low nucleotide and protein identity and differentiating genome level characteristics, there appears to be significant structural integrity and functional proximity between these proteins pointing towards their high significance. There is further need for comprehensive genomics and structural-functional studies to lead towards definitive conclusions regarding their criticality and that can eventually define their relevance to therapeutics development.

Keywords: COVID-19; SARS-CoV-2; SARS-CoV; Accessory Protein; ORF8; ORF8ab

1. Introduction

What started in early December 2019 as few cases of inexplicable pneumonia in Wuhan, China was officially named by World Health Organization (WHO) as Coronavirusdisease19 (COVID-19), and the International Committee on Taxonomy of Viruses (ICTV) classified causative agent as SARS-CoV-2 [1–3]. For its pan global nature of infections and deaths COVID-19 was declared a pandemic on 11th March, 2020 [4] and in around 07 months since the first reported case more than 15 million confirmed infected cases and above of 600,000 deaths all over the world have been recorded [5,6]. A standout characteristic of SARS-CoV-2 is the ability for asymptomatic transmission [7,8] and it is this unusual capacity to spread even during asymptomatic phases that has led to unprecedented measures to control its spread [9,10]. SARS-CoV-2 has proven to be an extraordinarily strong and lethal pathogen [11,12] impacting its primary site of infection i.e., nasal epithelium and lung [13,14] eliciting disproportional immune response with potential of leading to immune dysregulation [15]. Additionally, SARS-CoV-2 has been characterized by diverse disease spectrum spread over but not limited to gut [16], cardiovascular system [17], cutaneous system [18], central nervous system [19] with reports of autoimmune, autoinflammatory and multisystem inflammatory syndrome in children [20,21].

Similar to other coronaviruses SARS-CoV-2 is an enveloped virus characterized by a positive-sense, single-stranded RNA genome of approximately 30 kB that codes for six major open-reading frames (ORFs); ORF1a, ORF1b, spike (S), envelope (E), membrane (M) and nucleocapsid (N) [1,2]. SARS-CoV-2 spike protein [22–24], main protease [25,26], helicase and RNA-dependent RNA polymerase [27] have been classified as important targets for therapeutic intervention [28–31] for their roles in receptor identification, cell entry, viral replication, and transcription. In addition to these known factors of pathobiological and therapeutic importance, a vital coronavirus related characteristic is ‘accessory proteins’, the genes for which differ in genomic locations (**Figure 1A**), number and nature between coronavirus groups and have been also termed as ‘group-specific genes’ [32,33]. Although considered to be dispensable for viral replication and growth, their presence and maintenance within genomes has led to huge interest in understanding their significance to coronavirus life cycle and virulence [32–34]. Given the challenges that both COVID-19 and SARS-CoV-2 have posed, studies involving these pathogen specific proteins can therefore enhance understanding of its pathobiology and translate into new opportunities and targets for design of antiviral therapeutics.

SARS-CoV-2 genome is believed to harbor 6-9 accessory proteins with verified transcription regulatory sequence (TRS) identified upstream of ORFs 3, 6, 7, 8, and 10 [1] (**Table 1**). ORF8 is an accessory protein that is not shared by all members of subgenus sarbecovirus and it was the presence and location of ORF8 in SARS-CoV-2 genome that led to classification of SARS-CoV-2 genome with that of SARS-CoV [1]. SARS-CoV-2 Clade S subtype characteristics include the marker variant based on T28144C mutation leading to L84S change in ORF8 protein sequence [35], a positive selection that has resulted in the divergence of a separate phylogenetic group [36]. These accessory proteins offer

functional flexibility to coronaviruses and accordingly are subject to alterations depending upon the condition they are expressed during viral life cycle. SARS-CoV ORF8 homolog, ORF8ab from the closest human pathogen was accompanied by 29-nt nucleotide deletion during mid and late phase of the epidemic leading to two truncated and functional proteins ORF8a and ORF8b [37]. Irrespective of the debate around precise role of this deletion, biochemical and functional characterization of ORF8ab and its truncated counterparts has been pursued and is an ongoing process. The functional implications of SARS-CoV-2 ORF8 have already garnered attention and initial reports predict it to be an important component of immune surveillance machinery [38,39]. Keeping in mind the capacity of coronaviruses to periodically infect human population, it becomes imperative to view them in a holistic manner and study all aspects related to their life cycle and pathogenicity. This can potentially lead to better understanding of their evolution, host-to-human transmission processes, and pathobiology that will help development of better combating strategies as well as guide the course of future studies. In this context we analyze certain genome-based features of ORF8 and undertake a comparative evaluation of ORF8 and SARS-CoV ORF8ab genome stability, evolutionary origin, and protein characteristics.

Table 1. SARS-CoV-2 Accessory proteins. Genomic coordinates of ORFs along with their gene and protein length. Also mentioned is the transcription regulatory sequence (TRS) location for those accessory proteins for which it has been identified upstream of the start codon

ORF	Coordinates	Gene Length (nt)	CDS Length (nt)	TRS Location
3a	25,393-26,220	828	275	25,379
3b	25,765-26,220	456	151	
6	27,202-27,387	186	61	27,035
7a	27,394-27,759	366	121	27,382
7b	27,756-27,887	132	43	
8	27,894-28,259	366	121	27,882
9a	28,284-28,577	294	97	
9b	28,734-28,955	222	73	
10	29,558-29,764	117	38	29,528

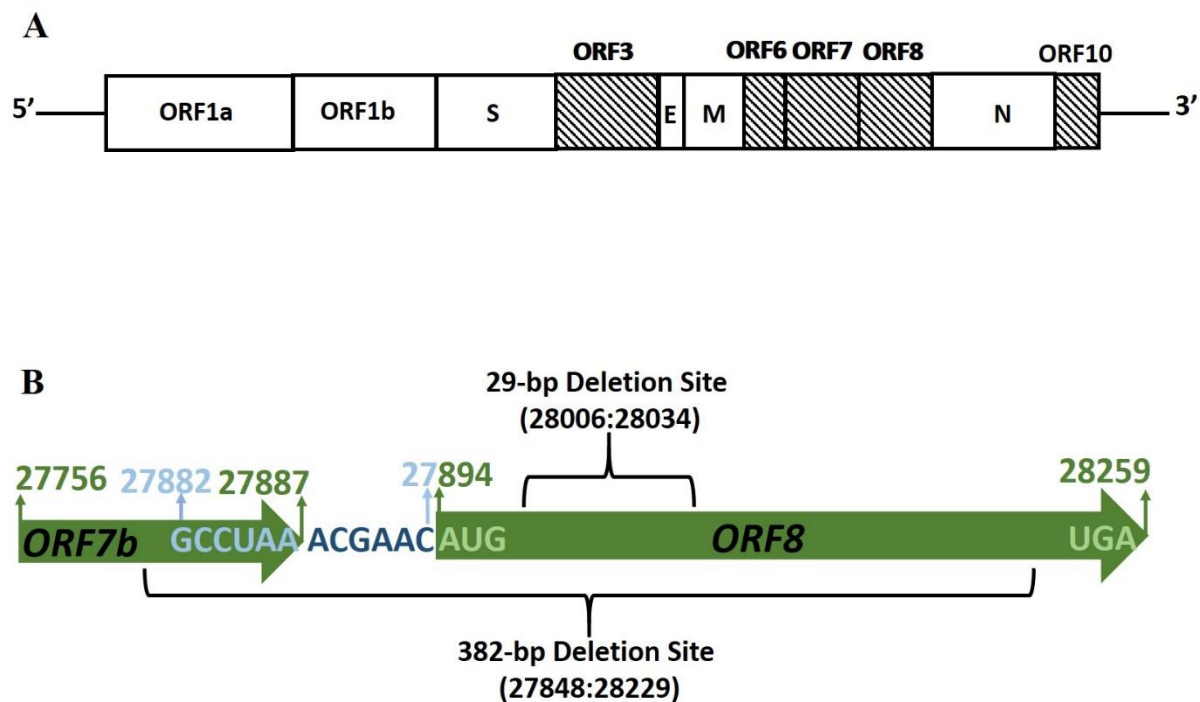


Figure 1. A. SARS-CoV-2 genome organization: ORFs with a verified transcription regulatory sequence (TRS). **B. SARS-CoV-2 Subgenomic mRNA8 region:** ORF regions and boundaries are highlighted in green. ORF8 transcription regulatory sequence (TRS) (27882:27894) is highlighted in blue and core TRS sequence in dark blue. Both the reported 482-bp deletion site and the predicted 29-bp deletion site are depicted.

2. ORF8 subgenomic mRNA8 stability

2.1 Genome deletions in SARS-CoV ORF8ab: During the SARS-CoV epidemic in 2003, subgenomic mRNA8 from animal sources and early human isolates coded for a full-length ORF8ab protein. An interesting and striking feature observed in human-to-human transmission during peak of this epidemic was a 29 nucleotide deletion (**Figure S1**) towards 5' region of ORF8, splitting it into two unequal ORFs; a smaller ORF8a and a long ORF8b that finally code for 39 and 84 amino acid long polypeptides respectively [40–42]. During late stages of this SARS epidemic, even larger deletions of 82nt and 415nt were identified in some virus clusters from human isolates, that lead to disruption of a putative ORF9 and eliminations of ORFs 10 and 11 [42,43]. Genetic and functional importance of 29nt deletion in SARS-CoV ORF8ab has been a matter of intense debate as to whether this is a case of genomic instability or adaptive evolution and has been hypothesized to contribute to zoonotic transition and favor human adaptation [33,41,42]. Contribution to human adaptation has been attributed to the functional implications of this deletion that lead to development of proteins with new functions relevant to later stages of a viral epidemic [41,42]. A contrary view to this functional implication for 29-nucleotide has been proposed, to be based in founder effect that has permitted SARS-CoV survival despite reduced fitness and is not somewhat of a generalized role in SARS-CoV

host adaptation [44]. Irrespective of the rationale behind this deletion, its existence in the SARS-CoV ORF8ab is a confirmed fact, but that has not deterred exploration of the functional significance (Section 5.1) of ORF8ab and its truncated versions ORF8a and ORF8b.

2.2 Genome deletions in SARS-CoV-2: COVID-19 has spread at an extremely fast rate in terms of both space and time. During a fast spreading and full-blown pandemic like COVID-19, there is possibility that an extremely large number of virions are generated and higher the number of these virions, more chances for genome related events like mutations and deletions. COVID-19 has spread at a rapid pace and almost all over the globe, therefore during such a fast spreading and full-blown pandemic expectedly an extremely large number of virions are generated and higher the number of these virions getting produced, more chances for genome related events like mutations and deletions. Three deletion events in the SARS-CoV-2 genome have been reported so far. Deletions in the region spanning ORFs 7a and 7b were detected during a surveillance program in Arizona, USA from among the samples screened in week of March 16-19, 2020 [45]. This 81 nucleotide deletion, leading to a 27 amino-acid in-frame deletion in one of the genomes AZ-ASU2923, maps to the putative signal peptide and first two beta strands in the protein structure [45]. A 382-nt deletion (27848:28229) has been reported from three hospitalized patients in Singapore in the genomic region spanning ORF7b and this deletion includes the transcriptional regulator sequence (TRS) of ORF8 [46]. A similar deletion was recorded in the genome (CGMH-CGU-02) of a sample isolated from a patient in Taiwan on February 4, 2020 and the patient had returned from Wuhan, China a day earlier [47]. Phylogenetic analysis of the cluster in which the 382-nucleotide deletion was reported from Singapore predicts possibility of a single source [46] and by nature of closeness this deletion can be considered to be of same origin as one from Taiwan. Therefore, these cases of 382-nt deletion having origin in isolates obtained from patients that had returned from Wuhan points towards their existence in earlier phase of outbreak. It is pertinent to note that the genome deletion events during viral life cycle and evolution are not usually random and are believed to play a role in helping viral genomes get rid of accumulated deleterious genome changes [48]. Because the deletions that have been reported so far from SARS-CoV-2 genomes occurred relatively in early phase of this pandemic, therefore this cause for these deletions can be ruled out. This argument is strengthened by recent reports about functional importance of ORF8 in immune modulation directing more towards indispensability of this protein [38,39]. Further discussions are in the ensuing section.

2.3 Genomic Stability estimation of SARS-CoV-2 ORF8: SARS-CoV-2 has proved to be a highly efficient pathogen in its capacity to be able to infect people all over the world and at a very rapid pace. Genome sequencing of large number of isolates from all over the globe has been made possible by vast progress made in next generation sequencing (NGS) via development of high-throughput platforms and other automations. Between the period between first report of SARS-CoV-2 genome sequence in December, 2019 and 30th June, 2020 data for around 30000 sequences in GISAID [35] and 7000 sequences in NCBI Virus [49] databases have been recorded. Keeping the enormity of this

genomic data in mind, it is plausible to believe that the deletions detected and reported so far do not provide a fair estimation of overall picture.

Keeping this in view we tried to run our own analysis targeting the sequences from both data repositories. For this purpose, representative sequence datasets named as NCBI Dataset (**Table S1**) and GISAID Dataset (**Table S2**) were utilized with NCBI Dataset containing 104 sequences and GISAID containing 177 sequences. These representative sets comprise of sequences reported between the period between December,2019 and end of June,2020. To achieve a global representation, genome sequences from Africa, Asia, Europe, North America, Oceania, and South America were included with a fair representation of the countries where high infection rates of COVID-19 were reported.

ORF8 gene is predicted to be 366 nucleotides (Genomic Region 27894:28259) in length and is preceded by a transcription regulatory sequence (TRS) on 5' end without any gap with initiation codon. The reported 382-nt deletion corresponds to genomic location 27848:28229, and the predicted 29-nucleotide deletion site on basis of similarity with ORF8ab can be mapped to 28006:28034. We therefore focused on the genomic region 27800:28300 that covers both the full-length ORF8 as well as the reported 382-nt deletion region (**Figure 1B**). We focused on stability of this genomic region to identify any deletion events, detected by analyzing the multiple sequence alignment (MSA) of the two representative sequence datasets. For the NCBI Dataset, multiple sequence alignment was straightway performed utilizing the online available tools on the NCBI Virus portal. The alignment file was viewed in the NCBI Multiple Sequence Alignment Viewer 1.15.0 focusing on the region 27800:28300 (**Figure 2A, Figure S2**). Whole genome multiple sequence alignment for the GISAID Dataset were performed utilizing MUSCLE [50]. Similar approach for visualization and analysis was utilized as for the NCBI Dataset utilizing the NCBI Multiple Sequence Alignment Viewer 1.15.0 (**Figure 2B, Figure S3**). In these sequence datasets subjected to multiple sequence alignment analyses, we did not identify any deletions other than the ones that have already been reported and an intact full-length ORF8 including expected 29-nucleotide deletion region of 28006-28034 can be seen. A recent comparative analysis of SARS-CoV-2, SARS-CoV, and other SARSr-CoV genomes has also led to identification of deletion vulnerable 430bp region in the SARS-CoV-2 ORF8 identical to SARS-CoV region but without detection of any deletion [51]. The unbiased spatio-temporal coverage that is provided by our two sequence datasets consisting of 381 representative genome sequences affords reasonable input to estimate genome stability. It looks reasonable to conclude that at this point SARS-CoV-2 subgenomic mRNA8 has to a large extent stayed stable and potentially codes for functional full length ORF8 protein.

The deletion events so far reported in the SARS-CoV-2 genome are limited to extremely small number of clusters and comparing to its global vertical and horizontal spread, can be presumed to be insignificant. Nonetheless, there has to be some reason behind their existence, a probable justification could be 'founder effect'. Founder effect is a phenomenon through which a genetically altered population of 'founders' is randomly selected through transmission bottlenecks and these founders retain the capacity to reproduce [52–54]. Founder effect as reason for genetic diversification has been

reported in many viruses and pathogens [55–57] including SARS-CoV [44], and in HIV has been an immense source of polymorphisms and evolution [55]. In a study evaluating distribution of mutations in SARS-CoV-2 genomes, clonal nature of mutations in different geographical regions was detected leading to assignment of founder effect as cause [59]. In the initial phases of endemic viral infections, viruses are under a constant selection pressure and certain lineages can eventually face extinction because of the competition with their more reproductively capable counterparts. However, during this initial period slightly deleterious mutations can continue to proliferate and their existence can be expected in early stages when the virus has not fully adapted to host environment [53]. The deletions [45–47] in SARS-CoV-2 genome have been reported in very early phases of this pandemic and from three specific locations, out of which origin for two, Taiwan and Singapore can be traced to Wuhan. The third deletion event has been reported from a specific location, Arizona in USA. Taken together, these deletions can be attributed to genetic events particular to these clusters and in absence of further reports about deletions looks like were specific to that period of time. In addition the number of isolates in these cases is low; three from Singapore [46] and one each from USA [45] and Taiwan [47], and with no further reports of additional deletion events from these geographic locations or anywhere else, it will be more plausible to conclude that SARS-CoV-2 genome has not undergone any effective genomic deletion events that can have an impact on its virulence and pathogenicity. Nevertheless, in spite of propensity for deletions in these genomic regions, exploration of functional importance of proteins they code for should not be compromised, especially keeping in view the ability of coronaviruses to repeatedly infect human population and with a SARS-CoV-2 like severity.

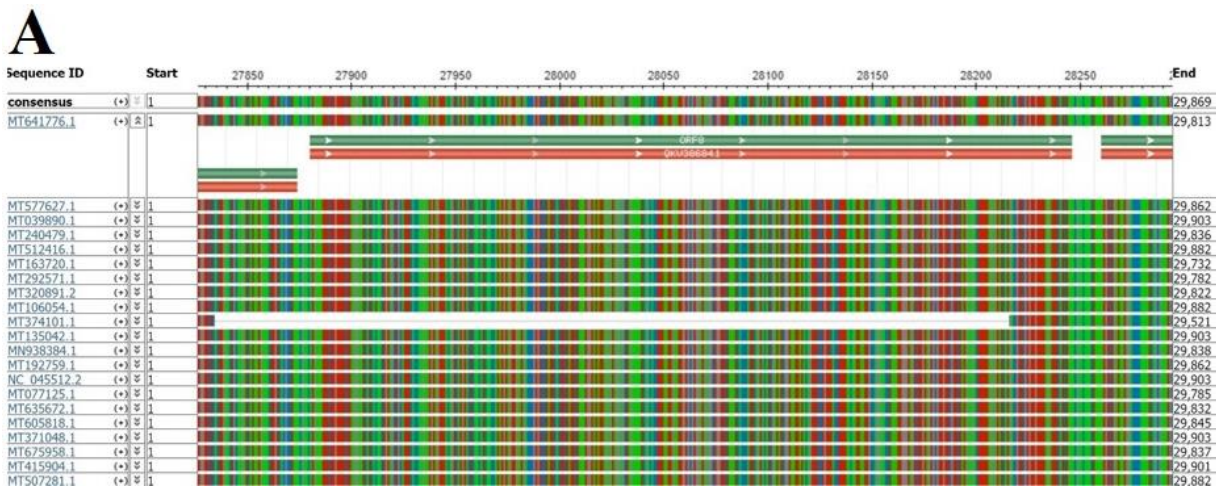


Figure 2A. Subgenomic mRNA8 region stability for NCBI Dataset. Representative figure for multiple sequence alignment. Intact ORF8 coding region is detected in all sequences other than for isolate MT374101.1 which is an already known deletion at the region 27848:28229. (Details in Figure S2)



Figure 2B. Subgenomic mRNA8 region stability for GISAID Dataset. Representative figure for multiple sequence alignment. Deletions can be visualized in already reported isolate sequences 419001 and 418996 at 27865:28246. The variation in deletion location from NCBI dataset is because of sequence alignment adjustment. NCBI dataset number represent the true numbers. (Details in Figure S3)

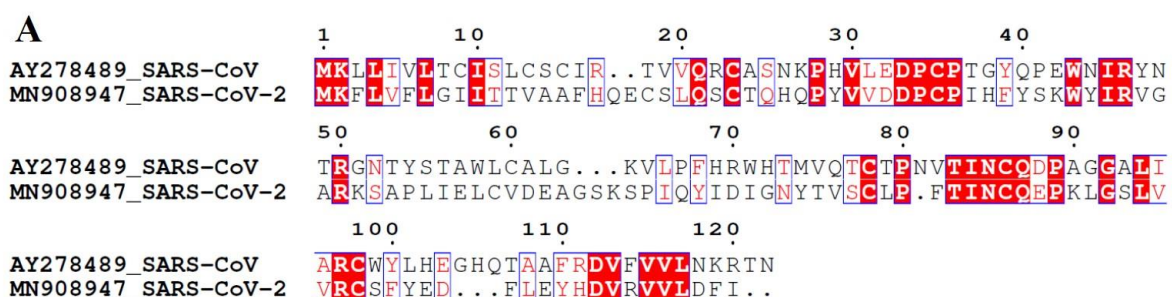
3. ORF8 protein origin

Coronaviruses belong to Order; *Nidovirales*, Family; *Coronaviridae*, Subfamily; *Coronavirinae*, and are ordered into four genera, Alpha, Beta, Gamma, and Delta [60]. SARS-CoV-2 along with the two recent coronaviruses that infected humans, Severe Acute Respiratory Syndrome-CoV (SARS-CoV), and Middle East Respiratory Syndrome CoV (MERS-CoV) belong to genus Betacoronavirus that is further classified into lineages A to D [42,61,62]. Bats are recognized as important reservoir of several emerging viruses that includes alphacoronaviruses and betacoronavirus lineages B, C, and D and, approximately 200 coronaviruses with bat origin have been identified with these coronaviruses regarded as major gene source for mammalian coronaviruses [63–65]. These viruses being RNA viruses are known to be prolific evolvers for their high rates of replication and mutation, coming from infidelity of their RNA-dependent RNA polymerase (RdRp) and that allows their fast movement through sequence space. In addition to this capacity to generate variations through mutations, these viruses also hold the capacity to exchange genetic material. Exchange of genetic material can occur through two distinct but not completely exclusive approaches; reassortment and recombination [48]. Reassortment is a trait associated with multipartite viruses like influenza A and involves swapping of discrete RNA molecule(s) from the segmented viral genome. Recombination on other hand, that can occur both in segmented and unsegmented viruses involves introduction of

‘donor’ nucleotide sequence into an ‘acceptor’ RNA molecule that can potentially contain genetic information from multiple sources.

3.1 SARS-CoV ORF8ab bat origin: SARS-CoV ORF8ab is believed to have originated from potential recombination events between betacoronaviruses from greater horse-shoe bats *Rhinolophus ferrumequinum* (SARSr-Rf-BatCoV) and Chinese horseshoe bats *Rhinolophus sinicus* (SARSr-Rs-BatCoV) [62]. SARSr-Rs-BatCoV and SARSr-Rf-BatCoV both share a remarkably high genome identity, 95% and 93% respectively with human/civet coronaviruses. ORF8ab amino acid identities do not reflect the same picture as is observed in genome identities; only SARSr-Rf-BatCoVs share 80.4-81.3% identity, while as SARSr-Rs-BatCoVs share only 32.2-33%. On basis of high genome identity between SARSr-Rf-BatCoVs, SARSr-Rs-BatCoVs and human/civet coronaviruses and high identity in ORF8ab between SARSr-Rf-BatCoVs and human/civet coronaviruses the authors came to conclusion that that the ancestor of civet SARSr-CoVs acquired its ORF8ab from SARSr-Rf-BatCoVs through its recombination with SARSr-Rs-BatCoVs. This was verified by identification of potential recombination sites between SARSr-Rf-BatCoVs and SARSr-Rs-BatCoVs around the ORF8ab region, leading to belief that civet SARSr-CoV SZ3 evolved with its ORF8ab acquired from SARSr-Rf-BatCoVs [62]. In a comprehensive study exploring coronaviruses from bats and analyzing their genetic variations with human/civet coronaviruses, phylogenetic clustering of SARSr-Rf-BatCoV ORF8ab phylogenetic clustering with human/civet SARSr-CoVs was observed [66]. Bat origin of ORF8ab specifically *Rhinolophus sinicus* has also been reported with direct identification of ORF8ab homologs in “SARS-like CoVs” (SL-CoVs) in these bats [67]. Thus, it can be concluded with confidence that SARS-CoV ORF8ab has primarily originated from bat coronaviruses. [67]

SARS-CoV ORF8ab and SARS-CoV-ORF8 share low nucleotide (26%) and protein (20%) identities (Figure 3A), thus making straightforward interpretations about evolutionary and phylogenetic relationships difficult. Likewise, this molecular identity makes it important to analyze the origin of SARS-CoV-2 ORF8 protein.



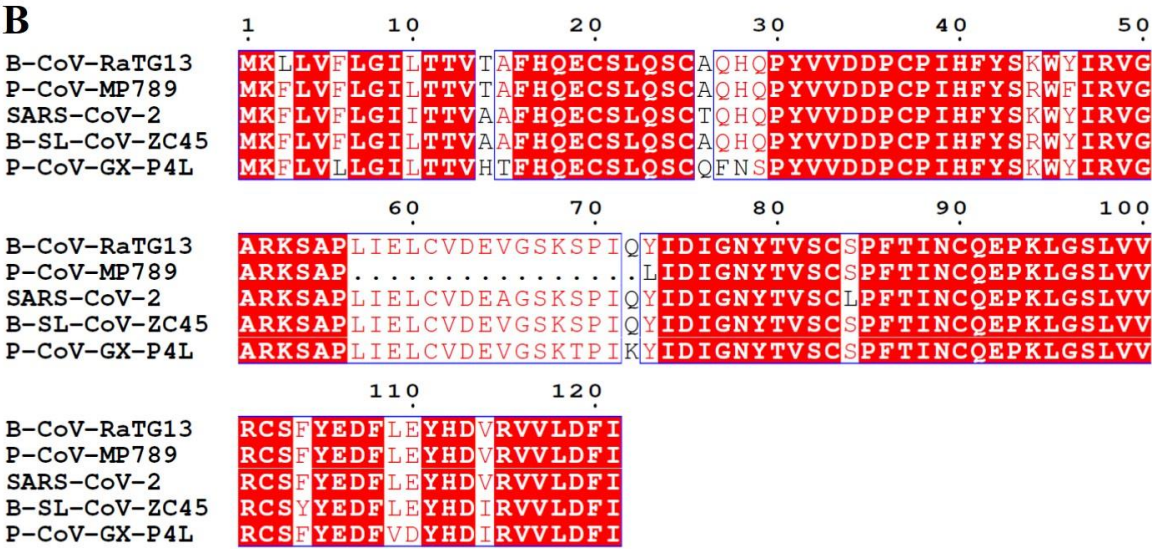


Figure 3. Protein sequence alignments of human SARS-CoV-2 ORF8 **A.** Alignment with ORF8ab (AAP51236.1) of SARS-CoV **B.** Multiple sequence alignment with the high identity (80-100%) proteins identified from blastp search against NCBI non-redundant protein sequences (nr) database. Sequence have been labelled according to isolate and predicted proteins corresponding to these isolates are non-structural protein (NS8) for Bat-CoV-RaTG13, a hypothetical protein for Bat-SL-CoVZC45, and ORF8 for two pangolin isolates.

3.2 SARS-CoV-2 ORF8 homologous proteins: As there is no published literature regarding ORF8 origin or about its homologous proteins, we therefore searched multiple genome and protein data repositories in pursuit of identifying its homologs. Using BLAST [68] searches, four proteins with identities ranging between 80-100% (**Figure 3B**) were identified after searching NCBI non-redundant protein sequences (nr) database, UniProtKB and Protein Databank (PDB). These four proteins were shortlisted for reason they provided stringency in terms of a 100% coverage against input SARS-CoV-2 ORF8 amino acid sequence (**Table 2**). The first two high identity homologs are from bat coronaviruses Bat-CoV-Ra TG13 and and Bat-SL-CoVZC45, among them there is remarkably high homology (95%) with BatCoVraTG13 non-structural protein 8 (NS8). This high homology with RaTG13 non-structural protein 8 is not surprising because it is the closest relative of SARS-CoV-2 with an overall nucleotide identity of around 93-95% [1,2]. Protein level identities between other SARS-CoV-2 and RaTG13 proteins have been recorded, for example, the SARS-CoV-2 S gene coding for receptor binding spike protein shares 93.1% nucleotide and 98% protein identity and this potentially affords some differentiating characteristics to SARS-CoV-2 spike protein [2,69]. In spite this close homology between bat and human infecting coronaviruses and the knowledge about bat source of coronaviruses, direct bat-to-human transmission has always been ruled out. In case of SARS [41] and MERS [61] these transmissions are believed to happen through zoonotic routes of civets and dromedary camels respectively. The fact that these animals play role either as intermediate or amplifying host has led to speculations that intense genome modification events within bats, and

intermediate hosts contribute to animal-to-human transmission and potentially to the virulence of these viruses [66,70]. Pangolins have received a considerable attention as an intermediate host for SARS-CoV-2 infection for the reasons that coronaviruses isolated from pangolins have significant genome level identity with SARS-CoV-2 genome as well as with bat coronaviruses that have high similarity with SARS-CoV-2 [71–74]. Therefore, it is not a surprise that the other two high identity homologs of ORF8 detected in our searches belong to pangolin coronaviruses; ORF8 of 81%, 86% with Pangolin-CoV-GX-P4L shares 81% identity while Pangolin-CoV-MP789 is 86% identical (Table 2). High amino acid identities with pangolin coronavirus homologs have been reported in case of other SARS-CoV-2 proteins also. A recently reported Malayan pangolin coronavirus isolate shares amino acid identities of 100% in envelope (E), 98.2% in main proteinase (M), 96.7% in nucleocapsid (N) and 90.4% spike (S) proteins, with the receptor-binding domains (RBD) of spike protein being almost identical carrying only a single amino-acid difference [71,72,75]. Thus, ORF8 behavior in terms of its genomic and protein identity is like its other counterparts from SARS-CoV-2 that show a strong identity with both bat and pangolin coronaviruses.

Table 2. High sequence identity homologs of ORF 8b protein.

Protein	Accession	Length [Coverage]	Percent Identity	Host	Isolate	Genome Accession	Genome Identity
Non- structural [NS8]	QHR63307.1	121[100]	95.04	Bat	CoV-Ra TG13	MN996532	95.98
Hypothetical	AVP78037.1	121[100]	94.21	Bat	SL- CoVZC45	MG772933	84.69
ORF8	QIA48620.1	121[100]	87.60	Pangolin	PCoV- GX-P4L	MT040333.1	80.17
ORF8	QIG55952.1	105[100]	81.82	Pangolin	CoV- Isolate MP789	MT121216.1	86.66

3.3 SARS-CoV-2 ORF8 evolutionary pathway: Genome level identities among SARS-CoV-2, bat and pangolin isolates identified in our ORF8 protein homolog searches are considerably high varying between 80–95% (Table 3). Some Pangolin coronavirus genomes have remarkably high homology (around 90%) with SARS-CoV-2 genome [71]; a recently identified Malayan pangolin, Pangolin-CoV-2020 shares 90.32% [73], Pangolin-CoV 91.02% [72], Pan_SL_CoV_GD 91.2% [76], and Pan_SL_CoV_GX 85.40% [76] identity with SARS-CoV-2. Therefore, a question arises whether ORF8 has descended directly from bat or pangolin coronavirus sources or some genome modification events have contributed towards its evolution. Coronavirus ORF8 region along with nsp3, ORF3 and

S are considered to be among rapidly evolving regions being flanked by recombination prone sequences [36,62,67], thus augmenting possible contribution of recombination in coronavirus and coronavirus element evolution. A recent publication has also reported presence of recombination breakpoints around ORF8 raising possibility of modular recombination occurring at both ends of ORF8 that are characterized by near identical nucleic acid sequences among SARS-CoVs and some bat CoVs [76]. They also observed that phylogenetic analysis in SARS-CoVs and analyzed bat coronaviruses based on the region around ORF8 led to their distinct and divergent clustering, pointing towards recombination around this region. In the initial phylogenetic analysis performed during its identification, the first SARS-CoV-2 isolate clustered with members of the subgenus Sarbecovirus but changed the topological position with respect to the gene used for phylogenetic estimation, leading to a conclusion that recombination has played a role in evolution of these coronaviruses [1].

With these facts and the close molecular identities between these ORF8 homologs harboring coronavirus isolates from bat and pangolin sources in mind, we tried to understand role of recombination in origin of SARS-CoV-2 ORF8. Recombination analysis involving the SARS-CoV-2 genome and four high identity genomes (**Table 3**) was performed using Simplot 3.5.15 [77] and RDP4 [78]. Bat-CoV-RaTG13 has highest sequence similarity with SARS-CoV-2 and also shares a close phylogenetic relationship and that is visible in the SimPlot genetic similarity plot where RaTG13 plot can be seen visibly separate from Bat-SL-CoVZC45, Pangolin-CoV-GX-789, and Pangolin-CoV-GX-P4L profiles (**Figure 4A**). Identification of potential recombinant regions was accomplished by sliding a 400-base pair (bp) window at a 50-bp step across the alignment using the Kimura (2-parameter) model. When SARS-CoV-2 was used as query with Bat-SL-CoVZC45, Pangolin-CoV-GX-789, and Pangolin-CoV-GX-P4L, several potential breakpoints can be identified (**Figure 4B**), but there is not a clear cut recombination crossover points around the subgenomic region coding for ORF8 (27894:28259). Utilizing RDP4 for analysis, total of 23 potential recombination cross-over points can be detected with the genomes involving different parental relations. Out of these 23, seven recombination events (**Table 4**) have the capacity to bring changes in SARS-CoV-2 genome, while as SARS-CoV-2 involvement as parent (**Table 5**) can be identified in 15 events, 09 out of which involve SARS-CoV-2 as the major parent. The results from both analyses suggest role of cross-species recombination is SARS-CoV-2 evolution, something that has been observed before also [76,79]. But similar to the SimPlot recombination analysis, none of the events identified in RDP4 analysis covers the ORF8 coding subgenomic region. These results are somewhat surprising in context of what is reported and expected about recombination around ORF8 (Starting lines of this paragraph), especially in view of the ORF8 protein identities with its homologs from bat and pangolin coronaviruses and the genome identities between these viruses. A simple and a plausible explanation could be that this analysis needs to be conducted with a larger sequence dataset that creates enough input for identifying and scoring these recombination events. At the same time, it needs to be kept under consideration that recombination is not the sole factor that contributes to evolution of

coronavirus proteins. As an example, the high amino acid similarity of SARS-CoV-2 spike protein receptor binding domain (RBD) with that of coronaviruses isolated from Pangolins in Guangdong, China has been proposed to be selectively mediated by convergent evolution [74]. Two amino acid substitutions 436Y and 427N in the RBD of SARS-CoV-2 spike protein are also present in SARS-CoV but not in the highly homologous spike protein of RaTG13, a potential adaptive convergent evolution in Sarbecovirus infecting humans [80]. Although too early to speculate but one is tempted to think about role of adaptive convergence in human infecting SARS-CoV ORF8s. The two proteins share low protein identity (20%), but in spite of that there is conservation of certain protein features (Section 4) and consequently functional characteristics (Section 5) among them. Even though SARS-CoV ORF8ab was accompanied by 29-nt deletion still the truncated versions retained their functions, and in fact functional diversity was speculated to be one of the reasons behind truncation. On the contrary as we saw in previous section that SARS-CoV-2 ORF8 has not undergone any significantly measurable deletion events, so its function as a full-length protein might be more important to its pathogenicity. Despite this potential variation in functional relevance of ORF8 to two pathogens, there is a significant overlap between them both in terms of functions and mechanisms behind them. A role for parallel or convergent changes is to create signals for adaptive evolution because it is highly unlikely that significant and complex characters can originate through multiple chances [81]. A deeper molecular level significant commonality between these proteins can only be identified from information about their structures which is lacking this time. In addition to established importance of structural importance, it is also important to help trace origin and evolutionary pathway of these proteins. Therefore, structural biology efforts must be made in earnest and should involve SARS-CoV-2 ORF8, its homologs from reservoir i.e., bat coronaviruses and intermediate host i.e., pangolin coronaviruses. The exploration of pangolin genomes is also being pursued vigorously and all this together will facilitate large studies that can help develop understanding of genetic processes involved in their evolution and unravelling of mechanisms behind acquirement of traits and functions that facilitate their animal-to-human transmission and virulence.

Table 3. Genome identities among human, bat, and pangolin coronavirus isolates in which SARS-CoV-2 ORF8 high identity homologs are detected

Subject	Target	Genome Identity
SARS-CoV-2	Bat-CoV-Ra TG13	95.98
	Bat-SL-CoVZC45	84.69
	Pangolin-CoV-GX-789	86.66
	Pangolin-CoV-GX-P4L	80.17
Bat-CoV-Ra TG13	Bat-SL-CoVZC45	85.3
	Pangolin-CoV-MP789	89.3
	Pangolin-CoV-GX-P4L	79.9

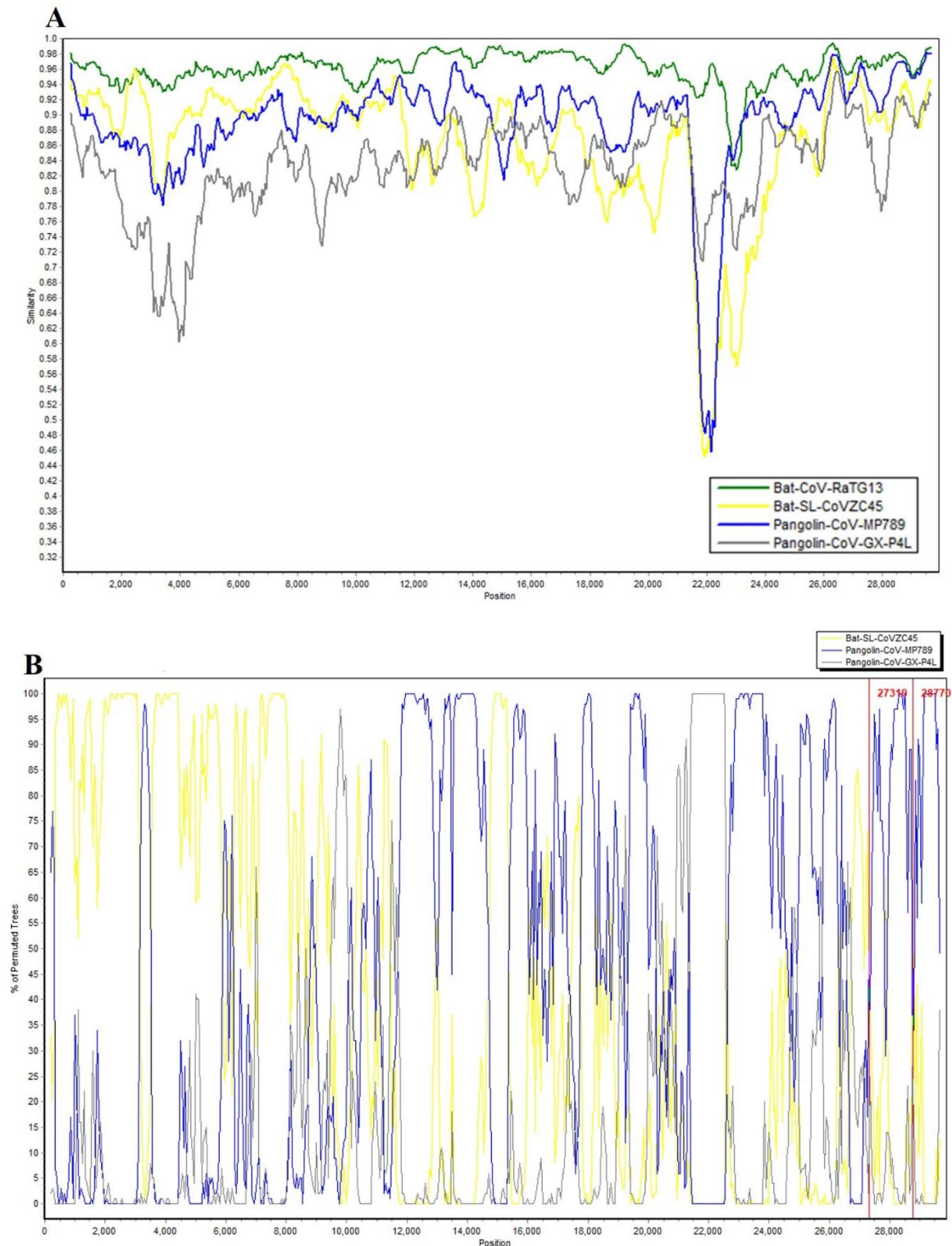


Figure 4. **A.** SimPlot genetic similarity plot between SARS-CoV-2 (MN908947.3) and sequences for bat and pangolin isolates for which ORF8 homologs show a high protein identity. **B.** BootScanning was conducted with Simplot version 3.5.1 using a 400–base pair (bp) window at a 50-bp step and the Kimura two-parameter model on a nucleotide alignment, generated with ClustalW. Highlighted region 27310–28770 harbors the ORF8.

Table 4. Recombination regions identified from RDP4 analysis in which SARS-CoV-2 is the recombinant sequence

No.	Begin	End	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	RDP P-value
1	927	1708	SARS-CoV-2	Bat-SL-CoVZC45	Pangolin-CoV-MP789	2.00*10 ⁻⁰⁶
2	1935	3194	SARS-CoV-2	Bat-SL-CoVZC45	Pangolin-CoV-MP789	3.05*10 ⁻¹¹
3	3664	4363	SARS-CoV-2	Bat-SL-CoVZC45	Pangolin-CoV-MP789	3.84*10 ⁻⁰⁹
4	22874	23092	SARS-CoV-2	Pangolin-CoV-GX-P4L	Bat-SL-CoVZC45	2.52*10 ⁻⁰²
5	23156	23306	SARS-CoV-2	Pangolin-CoV-GX-P4L	Bat-SL-CoVZC45	5.09*10 ⁻⁰³
6	23898	24248	Bat-CoV-RaTG13 SARS-CoV-2	Pangolin-CoV-GX-P4L	Bat-SL-CoVZC45	1.47*10 ⁻⁰³
7	6649	6833	Bat-CoV-RaTG13 SARS-CoV-2	Bat-SL-CoVZC45	Pangolin-CoV-MP789	3.19*10 ⁻⁰²

Table 5. Recombination regions identified from RDP4 analysis in which SARS-CoV-2 is involved either as major parent or minor parent.

Number	Begin	End	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	RDP P-Value
1	380	11623	Pangolin-CoV-GX-P4L	Unknown (Bat-SL-CoVZC45)	SARS-CoV-2	1.27*10 ⁻⁵⁷
2	7054	8258	Pangolin-CoV-MP789	Unknown (Bat-SL-CoVZC45)	SARS-CoV-2	1.36*10 ⁻¹⁰
3	9558	9947	Pangolin-CoV-MP789	Bat-SL-CoVZC45	SARS-CoV-2	2.20*10 ⁻⁰²
4	14611	15451	Pangolin-CoV-MP789	Unknown (Bat-SL-CoVZC45)	SARS-CoV-2	8.00*10 ⁻¹³
5	17813	18698	Bat-SL-CoVZC45	Unknown (Pangolin- CoV-GX-P4L)	SARS-CoV-2	9.99*10 ⁻⁰³
6	19847	19963	Pangolin-CoV-MP789	Pangolin-CoV-GX-P4L	SARS-CoV-2	5.09*10 ⁻⁰⁹
7	21563	21904	Pangolin-CoV-MP789	Unknown (Pangolin- CoV-GX-P4L)	SARS-CoV-2	1.34*10 ⁻⁰⁹
8	21914	22474	Pangolin-CoV-MP789	Unknown (Pangolin- CoV-GX-P4L)	SARS-CoV-2	3.73*10 ⁻²¹
9	22850	23094	Bat-CoV-RaTG13	Unknown (Pangolin- CoV-MP789)	SARS-CoV-2	1.70*10 ⁻¹⁶
10	4816	5953	Pangolin-CoV-MP789	Unknown (Bat-SL-CoVZC45)	Bat-CoV-RaTG13	1.25*10 ⁻⁰⁹
11	14042	14607	Bat-SL-CoVZC45	Unknown (Pangolin- CoV-GX-P4L)	SARS-CoV-2 Pangolin-CoV-MP789	2.13*10 ⁻⁰⁴
12	16028	16399	Bat-SL-CoVZC45	Unknown (Pangolin- CoV-GX-P4L)	SARS-CoV-2 Bat-CoV-RaTG13	2.07*10 ⁻⁰⁴
13	21187	22368	Bat-SL-CoVZC45	Pangolin-CoV-MP789	SARS-CoV-2 Bat-CoV-RaTG13	1.69*10 ⁻⁴⁵
14	20015	20591	Bat-SL-CoVZC45	Unknown (Pangolin- CoV-GX-P4L)	SARS-CoV-2 Bat-CoV-RaTG13	1.42*10 ⁻⁰⁷
15	22472	22792	Pangolin-CoV-GX-P4L	Bat-CoV-RaTG13	SARS-CoV-2 Unknown (Bat-SL- CoVZC45)	3.91*10 ⁻⁰⁴
				SARS-CoV-2		

4. Conserved features of SARS-CoV-2 ORF8 and SARS-CoV ORF8ab proteins

Knowledge of protein tertiary structure is critical to understanding of their functions as well to the understanding of their mechanisms of action. Atomic resolution structural details lead to molecular level information about unique protein features that can be exploited in therapeutics development. Keeping in mind the importance of protein structures, substantial efforts based in structural biology and computational protein modelling have been dedicated towards obtaining structural information about SARS-CoV-2 proteins. ORF8 has proved an intractable protein to computational biology based modeling approaches as neither templated-based homology modeling approaches are applicable nor machine-learning based tools like I-TASSER [82], AlphaFold [83] or even a newly developed approach [84] specifically for SARS-CoV-2 proteins has met with success. A comparative analyses of different modeling approaches applied to SARS-CoV-2 proteins concluded an unreliability and inefficacy of protein modelling approaches in certain SARS-CoV-2 proteins including ORF8 [84]. Conclusively, computational structural biology approaches at this stage might not be a reliable approach to study ORF8, therefore emphasizing the need for such studies and in interim making it more prudent to rely on methods like knowledge-based ones. We utilized web-based resources that together contain a diverse suite of predictive tools and can be salvaged to reliably deduce information about protein structural and functional features. SARS-CoV ORF8ab have been extensively characterized for its biochemical features, although atomic structural level information is still not available. Comparison between information generated about ORF8 here with the available knowledge about ORF8ab can be considered to be reliable enough to build a good understanding of protein features and will guide in design of further deeper and comprehensive structure-functions studies.

4.1 Endoplasmic reticulum Residence: The full-length SARS-CoV ORF8ab was identified as a stable endoplasmic reticulum resident protein by virtue of its cleavable N-terminal signal sequence, that directs its transport to ER [37]. The EGFP tagged 8ab and 8a fusion proteins displayed a quite reticular pattern similar while as 8b-EGFP fusion protein was found to be distributed all over the cells. The difference in localization is potentially due to the loss of N-terminal signal peptide as a result of 29-nt nucleotide deletion. The ER localization of ORF8ab was further validated by fluorescence colocalization studies with ER marker calreticulin and ER residence by pulse-chase experiments. Partly contrary to this, ORF8ab was found to be membrane bound identified through detergent resistance and microsome association behavior similar to calnexin, an ER-resident membrane protein [85]. Incidentally similar to previously mentioned study [37], in these experiments also calreticulin was used as control to identify ER luminal behavior. The differences in outcomes from the two studies can be potentially either due to difference in the cellular systems used (OST7-1 and HeLa Cells) or the method of evaluation (Immunofluorescence and Detergent Resistance). Irrespective of whether ORF8ab is membrane bound or soluble protein, its ER residence can be considered with conformity.

Sequence analysis of SARS-CoV-2 ORF8 predicts an N-terminal signal peptide located within amino acids 2-16 with 5-13 stretch having high hydrophobicity, a feature resonant with 6-15 hydrophobic core signal sequence characteristic [86–88]. Endoplasmic reticulum in contrast to cytosol and other cellular organelles has a predominantly oxidative environment that provides essential ecosystem for oxidation, protein folding and protein quality control processes [89,90]. In terms of its ion concentration and redox potential, ER ecosystem very well matches with the extracellular environment where secreted and surface proteins work [90–92]. One of the important functions that endoplasmic reticulum performs by virtue of its oxidative environment is introduction of intra- or intermolecular disulfide bonds between unpaired cysteine residues of polypeptides [89,93]. The cysteine content of secretory pathway and mammalian extracellular proteins is higher than cytosolic proteins, correlating with the requirement for oxidative protein quality system afforded by endoplasmic reticulum [89]. SARS-CoV-ORF8ab protein is characterized by presence of 10 cysteine residues and exists as disulfide-linked homomultimeric complexes in endoplasmic reticulum [37]. ORF8 has seven cysteine residues at positions 20, 25, 37, 61, 83, 90 and 102 and can be expected to engage in disulfide bridge formation with two separate predictions forecasting three disulfide bonds. Keeping these facts in consideration it can be reasonably concluded that ORF8 like SARS-CoV ORF8ab is an ER resident protein with similarities spanning the presence of N-terminal signal peptide sequence and potential to form disulfide bonds. That ORF8 is an ER resident protein can be gauged from the fact that its host interactome identified in HEK-293T/17 cells significantly consists of endoplasmic reticulum resident proteins involved in pathways like protein quality control [28].

4.2 Conversed Glycosylation Site: Glycosylation is an important step in protein maturation and involves attachment of sugar moieties that contributes to both stability and solubility of a protein. SARS-CoV ORF8ab is characterized by presence of an N-glycosylation site structured around asparagine 81 with overall motif organization as Asn-Val-Thr [37,94]. N-linked glycosylation involves oligosaccharyl transferase (OST) complex facilitated attachment of core Glc₃Man₉GlcNAc₂ (3 glucoses, 9 mannoses and 2 N-acetylglucosamines) to the amino acid nitrogen atom, which is usually N4 of asparagine [95,96]. This interaction happens at typical sequence motifs with a design Asn-X-Thr/Ser/Cys, in the decreasing order of probability between Thr, Ser, and Cys and where X is a residue other than proline [97,98].

At least one N-glycosylation motif can be very well identified in ORF8 sequence with asparagine 78 being potential site of glycosylation and the motif as Asn-Tyr-Thr (**Figure 5**). The relatively similar locations, ORF8ab glycosylation site position at Asn-81 and ORF8 glycosylation site at Asn-78 can be construed as an indicator of potential structural conservation between these proteins. Protein glycosylation is an important aspect of viral biology and pathogenicity and is a known fact that enveloped viruses like coronaviruses very well use the host cell glycosylation machinery. Although viral protein glycosylation is relevant in functions like viral attachment, cell entry, assembly and exit, and viral spread, but the most important function that has been assigned to glycosylation is role in evasion of host immune system [99–101]. Cryo-EM and mass spectrometric

analysis of SARS-CoV-2 spike protein has led to identification of N-linked glycan occlusion of receptor binding sites, a feature observed in other viral glycoproteins with a purpose to conceal a structurally conserved and functionally important region of the spike protein [102]. Because ORF8 is not in all probability involved in functions related to cell entry or fusion, we therefore speculate that protein glycosylation has a role in its folding and stability function over a role in immune system circumventing. In fact, a structure stabilizing role has been detected for glycosylation in SARS-CoV ORF8ab, where in the glycosylation-defective 8ab mutant was found to be unstable similar to truncated protein 8b that has lost its glycosylation site by virtue of 29-nt deletion [94].

A unified picture (**Table 6**) that can be drawn of ORF8 and ORF8ab is that they are endoplasmic resident proteins transported there by virtue of an N-terminal hydrophobic signal peptide. They are further characterized by presence of a conserved N-glycosylation site potentially playing a role in their stabilization. The presence of cysteine residues with capacity and potential role to form disulfide bonds points towards higher level of structural organization and a greater functional role in the viral pathobiology.

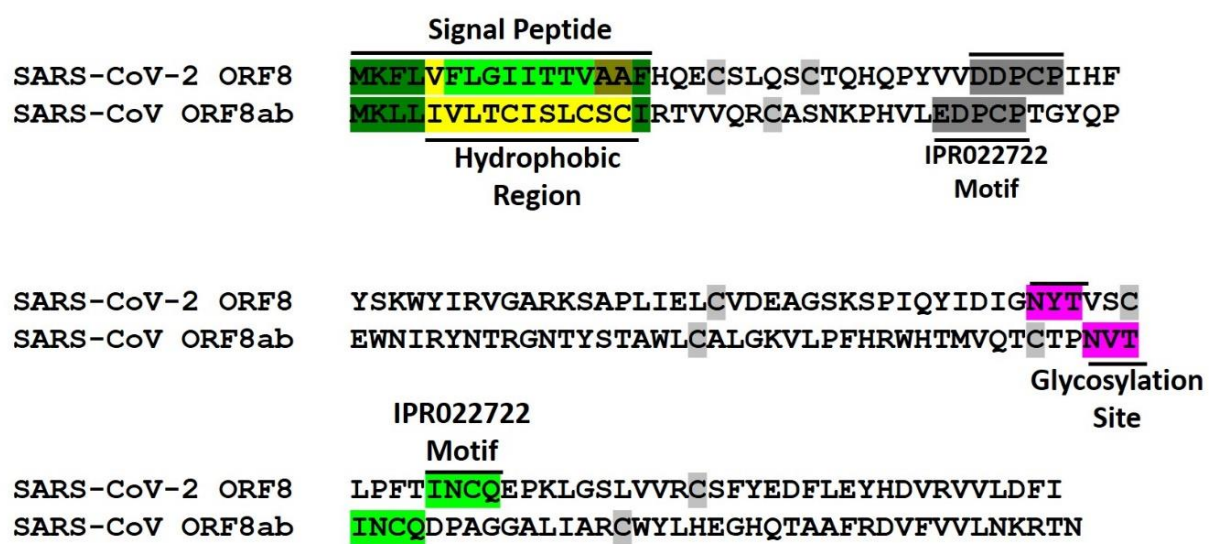


Figure 5. Protein characteristics of SARS-CoV-2 ORF8 and SARS-CoV-ORF8ab. Sequences are aligned to depict precise positions of sequence-based features and with respect to one another. INCQ and EDPCP are two conserved motifs in proteins belonging to InterPro family IPR022722. Cysteine residues have been highlighted in grey other than at position that fall within other conserved sites.

Table 6. Summary of SARS-CoV-2 ORF8 and SARS-CoV-ORF8ab features

CHARACTERISTIC	SARS-CoV ORF8ab	SARS-COV-2 ORF8
Nucleotide Identity		26%
Protein Identity		20%
Nucleotide Deletion	Yes	No (This Study)
Origin	Bat	Bat, Pangolin (This Study)
BIOCHEMICAL/STRUCTURAL FEATURES		
N-Terminal Peptide Sequence	Yes	Yes (Predicted in this Study)
N-Glycosylation Site	Yes	Yes (Identified in this Study)
Cysteine Residues	Yes	Yes
Localization	Endoplasmic Reticulum	Endoplasmic Reticulum (Predicted/proposed in this Study)
Protein Family Conserved Motifs	Yes	Yes
Macromolecular Interactions	Protein-protein	Protein-protein and Protein-DNA
FUNCTIONAL FEATURES		
Viral Replication	Yes	Not studied so far
Host Immune Modulation	Yes	Yes
Protein Quality Control	Yes	Yes

5. Functional landscape of SARS-CoV-2 ORF8 and SARS-CoV ORF8ab

5.1 SARS-CoV ORF8ab, ORF8a and ORF8b Functions: Despite confirmed deletion of a 29-nucleotide region, the functional importance of three translation products of SARS-CoV sub-genomic mRNA8, full-length ORF8ab, ORF8a and ORF8b has been a matter of significant interest. Multiple studies have been conducted to identify their roles in pathogenicity modulation, virus growth, replication, and host interactions.

5.1.1 Viral Replication: One of the highly studied and hotly contested functional importance of SARS-CoV ORF8ab and its truncated versions is the role in replication with results from different studies pointing towards both importance and lack of role in viral replication. While some studies have reported role for truncated ORFs 8a and 8b in viral replication [103,104], an intact ORF8ab [105] and ORF8b [106] were not found to impact replication efficiency, duration of replication, or SARS-CoV pathology. In a recent comprehensive study evaluating the impact of 29nt deletion on replication properties of SARS-CoV in relevant models of human respiratory tract infection and other organisms that included primate, *Rhinolophus* bat, cotton rat, goat and sheep cells, a 23-fold decrease in replication was observed [44]. The authors further assign the observed discrepancy (3-fold reduction) [105] in fold decrease in replication to the high MOIs used in this study, that are not close to doses at which natural infections occur. However, they have not dealt upon the mechanistic details of this ORF8ab mediated viral inhibition, but there have been other reports that proposed ORF8b mediated inhibition of replication through inhibition of the envelope (E) protein of SARS-CoV. 8b interaction with other SARS-CoV proteins like membrane (M), ORFs 3a and 7a, in addition to envelope (E) has been detected [107]. Reduction in E protein expression as a consequence of 8b overexpression was identified that acts through a post-translational mechanism and not through a mechanism that can lead to reduction in E gene expression. This down-regulation of E by 8b in a later study was found to play a negative role in viral replication involving a ubiquitin-independent proteasome pathway [106]. Ubiquitin independent nature of this down regulation was deduced through mutational analysis of 8b lysine residues that were found to be critical only for monoubiquitination. Ubiquitination potential of ORF8ab and 8b has been reported with these proteins reported to be post-translationally modified by ubiquitin with ability to form both mono- and polyubiquitinated species [94].

Inhibition of replication might seem counter-intuitive to the requirement for viral proliferation but one of the purposes of viral evolution is attainment of balance between effective transmission and evasion of host immune response. Consequently, the viral replication inhibitory effect of ORF8b has been ascribed to a replication modulation function with the 29-nt deletion contributing to fine-tuning of viral-replication [107], because ORF8a [103] has been reported to work contrary to 8b in a pro-replication role. Nonetheless, a conclusion that can be drawn is that subgenomic mRNA8 of SARS-CoV leads to translation products that play an essential role in viral replication and consequently in its pathogenicity.

5.1.2 Immune Modulation: The other important function of ORF8ab or its truncated counterparts that has garnered interest is role in immune modulation. This is not surprising as one of the primary strategies during viral infection involves overcoming host immune response. That SARS-CoV pathogenesis is to some extent controlled by immune signaling was deduced by association of SARS-CoV with anomalous cytokine and chemokine responses and expression of interferon stimulated genes (ISGs) in infected patients [108,109]. Multiple models of SARS-CoV infection have led to identification of critical components of innate signaling pathways that have protective role against SARS-CoV disease [108], and thus a viral response to circumvent this challenge is expected. Roles of

non-structural proteins like nsp1, nsp7, nsp14, nsp15, and nsp16, structural proteins like membrane, and nucleocapsid and accessory proteins like ORF3b, and ORF6 have been comprehensively elucidated [33,108]. ORF8ab and ORF8b were reported to work as IFN antagonists having a role in evasion of immune surveillance system by acting at one of the initial steps of the INF- β signaling cascade. Both ORF8ab and the truncated counterpart were found to physically interact with interferon regulatory factor 3 (IRF3) and leading to its degradation in ubiquitin proteasome dependent manner [109]. Ubiquitin proteasome system (UPS) mediated proteasomal degradation is a common strategy employed by many viruses in immune surveillance evasion by targeting host immune signaling and adaptor molecules like those involved in type I interferon (IFN) response and MHC class I antigen presentation [110]. Thus, one of the functions SARS-CoV subgenomic mRNA8 products are involved is in immune evasion through hijacking of host ubiquitin proteasome system (UPS). Other known mechanism by which a translation product of subgenomic mRNA8 has been reported to be involved in immune evasion is through formation of insoluble protein aggregates. Existence of protein aggregates post viral infections has been attributed to their usage as replication scaffolds and a protective mechanism against host induced protein degradation [111]. ORF8b forms insoluble intracellular aggregates in macrophages leading to activation of NLRP3 inflammasomes. This leads to activation of transcription factor EB (TFEB), the master regulator of autophagy and lysosome machinery and is facilitated by direct interaction of ORF8b with leucine rich repeat domain of NLRP3 [112]. These nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) are involved in activation of diverse signaling pathways through factors that lead to production of immune effectors like type I interferons (IFNs), interleukin-1 β (IL-1 β) and IL-18, while some members have been shown to inhibit NF- κ B and type I IFN-mediated signaling pathways, pointing towards role of NLRs in control of innate immunity [113,114]. Thus, like other viruses SARS-CoV involves strategies that allow it to evade immune surveillance system and ORF8ab and ORF8b form as essential component of this process.

5.1.3 Unfolded Protein Response (UPR) Modulation: A process level function that has been reported for ORF8ab is regulation of unfolded protein response (UPR). This happens through upregulation of synthesis of endogenous endoplasmic reticulum (ER) resident chaperones. Ectopic ORF8ab induced expression was observed at promoter of glucose regulated protein 78 (GRP78), one of the best characterized ER chaperones, and ORF8ab also led to upregulation of other ER-resident chaperones like GRP94 and CRT [85]. GRP78 also known as immunoglobulin heavy-chain binding protein (BiP) is an ER master controller that through interactions with PKR-like ER kinase (PERK), activating transcription factor 6 (ATF6) and the ER transmembrane protein kinase/endoribonuclease (IRE1) to ensure differential regulation of ER stress required during viral pathogenesis and replication [115]. ORF8ab mediated transcriptional activation of GRP78 was found through activation of ATF6, which binds to ER stress response element (ERSE) promoter element of ER chaperones and is an essential requirement for their transcriptional activation. This ATF6 activation by ORF8ab occurs through direct interaction between the two proteins in ER lumen leading to ATF6 movement into Golgi,

induction of its proteolytic cleavage and forward release of its N-terminal fragment to the nucleus for activation of expression of target genes of ER stress.

A common theme in the functionality of ORF8ab and its truncated counterparts is the capacity to bind proteins be they viral proteins like membrane (M), ORF3a, ORF7a, and envelope (E) [107] or the host proteins like interferon regulatory factor 3 (IRF3) [109]. These physical interactions with proteins lead to their degradation and eventual down-regulation thus impacting the processes they are involved in. The mechanistic spectrum through which these functions are achieved include apoptosis [103], proteasomal degradation [106], autophagy [112], and unfolded protein response [85]. Therefore, the overall picture that can be drawn of ORF8ab, ORF8a, and ORF8b as components of protein homeostasis. To firmly establish the existence and role of these functions more coordinated studies need to be undertaken that remove the potential variation and bias created from choice of cellular systems and experimental methodologies.

5.2 SARS-CoV-2 ORF8 Functions: In spite of short duration since start of COVID-19 pandemic, SARS-CoV-2 ORF8 has garnered reasonable attention and few studies related to its function have been conducted.

5.2.1 Immune Modulation: Overexpression of ORF8 led to significant down-regulation of MHC-I (HLA-A2) in 293T cells, human fetal colon cell line FHC, human bronchial epithelial cell line HBE, and human liver cell line Huh7 [38]. ORF8 mediated downregulation of MHC-I in 293T was ascribed to lysosomal degradation pathway with ORF8 binding of MHC-I leading to protection against cytotoxic T-lymphocyte (CTL) mediated lysis of SARS-CoV-2 infected cells. ORF8 strongly colocalized with Calnexin and LAMP1, ER and lysosomal proteins respectively alluding towards MHC-I degradation in endoplasmic reticulum via trafficking from ER to lysosome. Both endoplasmic reticulum associated degradation (ERAD) pathway and ubiquitination mediated degradation of MHC-I were ruled out from knock down experiments involving ERAD pathway proteins and absence of significant ubiquitination of MHC-I [38]. While this potentially offers a unique mechanism for evasion of immune surveillance and probable source of therapeutic intervention, the authors of this study as well as a follow up commentary warrant need for further detailed mechanistic studies [38,116]. In another study analyzing the SARS-CoV-2 response to host innate immunity ORF6, ORF8, and nucleocapsid protein were identified as potential inhibitors of type I interferon signaling pathway acting through inhibition at promoter sites and also by inhibiting interferon-stimulated response element (ISRE), but with a low activity for ORF8 [39]. A unique COVID-19 disease phenotype has been described from severe and critical patients, that consists of a highly impaired interferon (IFN) type I response that is characterized by low IFN- α production and activity and no IFN- β that was associated with persistent blood viral load [117]. Thus, ORF8 is apparently involved in one of the most important pathways of viral pathobiology and might be reason that the subgenomic mRNA8 region of SARS-CoV-2 has not undergone significant deletions. Although SARS-CoV ORF8ab and 8b were also reported to have similar function achieved by inhibition of interferon regulatory factor 3 (IRF3) [109] through a protein-protein interaction, a protein-DNA

interaction can be speculated in case of the SARS-CoV-2 ORF8 mediated interferon signaling pathway inhibition [38]. But this will be too premature to conclude and will therefore need validation and further studies that can uncover the mechanistic aspects. Nevertheless, there appears to be a reasonably high functional overlap between the two homologs.

5.2.2 Endoplasmic Reticulum Protein Quality Control: Another report that although was not purposed to study ORF8 functions directly, still provides some clues about its role. The purpose of this study was to evaluate the SARS-CoV-2 human protein-protein interactions within human cells (HEK293T/17) with a goal of identifying any potential therapeutic targets [28]. In this interactome analysis ORF8 interactions were detected with proteins involved in endoplasmic reticulum protein quality control, glycosylation, extra-cellular matrix organization (ECM), and glycosaminoglycan synthesis. With a rationale that host-virus interface processes and host targets are less prone to mutations, and can be therefore subject to therapeutic intervention, IVHR-19029 a known potent endoplasmic reticulum α -glucosidases I and II inhibitor was identified as a molecule of interest that can be targeted to ER Protein Quality Control for its member interactions with ORF8 [28]. This molecule has also been identified as a potent antiviral in other viral infections like hemorrhagic fever viruses [118,119].

Reading these results together with the role in MHC I downregulation and inhibition of interferon signaling, initial inferences can be made that SARS-CoV-2 functions involve impacting different viral or host-pathogen processes with interactions with other macromolecules like proteins and DNA. This macromolecule interacting nature of ORF8 at the protein level matches very well with that of SARS-CoV ORF8ab which interacts with both viral [107] and host proteins like interferon [109]. Development of a comprehensive and meaningful understanding of ORF8 necessitates requirement for large-scale system level studies like proteomics conducted in systematic and temporal manner that can help ascertain its stage specific role and analyze impact on other viral and cellular processes.

6. Conclusions:

SARS-CoV-2 is the seventh coronavirus to infect humans that primarily cause respiratory illnesses; NL-63, 229E, HKU1 and OC43 lead to mild illness, MERS-CoV and SARS-CoV cause more serious form of disease [120] while as SARS-CoV-2 has proved a different kind of challenge. Coronaviruses are significantly hosted in bats and because of them being RNA viruses hold a remarkable capacity to mutate and potentially lead to existence of large number of viral variants. This is exacerbated by their propensity for inter-species zoonotic transitions that potentially contributes to further complexity in their genomes. All these factors make them potent and formidable pathogens with capacity to infect and challenge for a long time, therefore, a considerable public health concern. Keeping this in consideration there need to be concerted efforts towards developing understanding of their pathobiology with the aim of identifying new targets, pathways, and approaches for therapeutics development. Most of current therapeutic strategies rely upon known proteins of pathobiological importance like spike protein, main protease, helicase and RNA-dependent RNA polymerase. The way COVID-19 is presenting itself has necessitated approaches like drug-

repurposing, and a relentless pursuit towards vaccine development. Keeping all these factors under consideration, it will be prudent to expand the base towards developing our understanding of coronavirus biology and pathogenicity with the hope that it can lead to development of new avenues for therapeutics development.

We have attempted here to comparatively study one of the accessory proteins pertaining to subgenomic mRNA8 from two recent human infecting coronavirus pathogens (**Table 6**). The two proteins have remarkably low identity at nucleotide (26%) and protein level (20%) probably due to their different origins; ORF8 looks to have originated from both bat and pangolin as compared to SARS-CoV ORF8ab that is of bat origin. At protein level both proteins are characterized by presence of an N-terminal hydrophobic signal peptide, a conserved N-glycosylation site, and enough cysteine residues with potential to form disulfide bonds, drawing their picture as structurally stable potential ER-resident proteins. There is functional overlap between these proteins with involvement in immune modulation that is probably accomplished through involvement in protein quality control. The purported functioning of ORF8 through endoplasmic reticulum associated degradation (ERAD) [28,38] and SARS-CoV ORF8ab in unfolded protein response (UPR) [85] deal with two different aspects of ER protein quality process, but the two processes are not completely independent [121,122], thus pointing to some overall process driven role for two less identical homologs. It could possibly be that these proteins belong to coronavirus specific protein family. To firmly establish this or their role in coronavirus pathobiology, it will need further comprehensive studies that involves approaches based in areas like genomics and evolution, systems-biology, proteomics, and structural biology. We believe that these studies should not be limited to known human pathogens only but should target other zoonotic viruses as well. Structure studies have an added advantage that they can identify unique domains or other structural motifs and can provide insights into mechanistic details that can be exploited in development of anti-viral therapies.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1,

Figure S1: 29-nt deletion in SARS-CoV ORF8ab

Figure S2: SVG File for Multiple Sequence Alignment of NCBI Sequence Dataset

Figure S3: SVG File for Multiple Sequence Alignment of GISAID Sequence Dataset

Table S1: NCBI Dataset Information

Table S2: GISAID Dataset Information

Material S1: Supplementary Methods

Author Contributions: Conceptualization, S.S.M., S.M., A.B., and T.S.K.; Methodology, S.S.M., M.R.; Sequence Download and Alignments, M.R, B.M.A., and N.S.G.; Bioinformatics Analysis, S.S.M., M.R., B.M.A., and N.S.G.; Writing—original draft preparation, S.S.M.; Writing—review and editing, S.S.M., S.M., B.M.A., A.B., and T.S.K.

Funding: This is King Abdullah International Medical Research Center, MNGHA, Riyadh project number RC20/222/R

Conflicts of Interest: “The authors declare no conflict of interest.”

References

1. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269, doi:10.1038/s41586-020-2008-3.
2. Zhou, P.; Yang, X. Lou; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273, doi:10.1038/s41586-020-2012-7.
3. On, C.S.G. of the I.C. on T. of V. The species Severe acute respiratory syndrome- related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **2020**, *5*, 536–544, doi:10.1038/s41564-020-0695-z.
4. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 Available online: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed on Jul 22, 2020).
5. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) Available online: <https://coronavirus.jhu.edu/map.html> (accessed on Jul 23, 2020).
6. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet. Infect. Dis.* **2020**, *3099*, 19–20, doi:10.1016/S1473-3099(20)30120-1.
7. Wölfel, R.; Corman, V.M.; Guggemos, W.; Seilmaier, M.; Zange, S.; Müller, M.A.; Niemeyer, D.; Jones, T.C.; Vollmar, P.; Rothe, C.; et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* **2020**, doi:10.1038/s41586-020-2196-x.
8. Liu, Y.; Yan, L.M.; Wan, L.; Xiang, T.X.; Le, A.; Liu, J.M.; Peiris, M.; Poon, L.L.M.; Zhang, W. Viral dynamics in mild and severe cases of COVID-19. *Lancet Infect. Dis.* **2020**, *2019*, 2019–2020, doi:10.1016/S1473-3099(20)30232-2.
9. Stephen M.Kissler, Christine Tedijanto, Edward Goldstein, Yonatan H.Grad, M.L. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **2020**, *21*, 1–9, doi:10.1016/j.solener.2019.02.027.
10. Jia, J.S.; Lu, X.; Yuan, Y.; Xu, G.; Jia, J.; Christakis, N.A. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **2020**, 1–11, doi:10.1038/s41586-020-2284-y.
11. Bradley, B.T.; Maioli, H.; Johnston, R.; Chaudhry, I.; Fink, S.L.; Xu, H.; Najafian, B.; Marshall, D.; Lacy, J.M.; Williams, T.; et al. Histopathology and Ultrastructural Findings of Fatal COVID-19 Infections. *medRxiv* **2020**, 2020.04.17.20058545, doi:10.1101/2020.04.17.20058545.
12. Gupta, A.; Madhavan, M. V.; Sehgal, K.; Nair, N.; Mahajan, S.; Sehrawat, T.S.; Bikdeli, B.; Ahluwalia, N.; Ausiello, J.C.; Wan, E.Y.; et al. Extrapulmonary manifestations of COVID-19. *Nat. Med.* **2020**, *26*, doi:10.1038/s41591-020-0968-3.
13. Liao, M.; Liu, Y.; Yuan, J.; Wen, Y.; Xu, G.; Zhao, J.; Cheng, L.; Li, J.; Wang, X.; Wang, F.; et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **2019**, doi:10.1038/s41591-020-0901-9.
14. Sungnak, W.; Huang, N.; Bécavin, C.; Berg, M.; Queen, R.; Litvinukova, M.; Talavera-López, C.; Maatz, H.; Reichart, D.; Sampaziotis, F.; et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **2020**, *26*, doi:10.1038/s41591-020-0868-6.
15. Hughes, R. Dysregulation of immune response in patients with COVID-19 in Wuhan, China

- Chuan. *J. Infect. Dis.* **2020**, doi:10.1017/CBO9781107415324.004.
16. Gu, J.; Han, B.; Wang, J. COVID-19: Gastrointestinal Manifestations and Potential Fecal–Oral Transmission. *Gastroenterology* **2020**, *158*, 1518–1519, doi:10.1053/j.gastro.2020.02.054.
 17. Endotypes, A.; Hariri, L.; Ph, D.; Hardin, C.C.; Ph, D. Covid-19, Angiogenesis and ARDS endotypes. *N. Engl. J. Med.* **2020**, 1–2, doi:10.1056/NEJMe2018629.
 18. Galvan Casas, C.; Catala, A.; Carretero Hernandez, G.; Rodriguez-Jimenez, P.; Fernandez Nieto, D.; Rodriguez-Villa Lario, A.; Navarro Fernandez, I.; Ruiz-Villaverde, R.; Falkenhain, D.; Llamas Velasco, M.; et al. Classification of the cutaneous manifestations of COVID-19: a rapid prospective nationwide consensus study in Spain with 375 cases. *Br. J. Dermatol.* **2020**, bjd.19163, doi:10.1111/bjd.19163.
 19. Cantuti-castelvetri, L.; Ojha, R.; Pedro, L.D.; Djannatian, M.; Franz, J.; Kuivanen, S.; Kallio, K.; Kaya, T.; Anastasina, M.; Joensuu, M.; et al. Neuropilin-1 facilitates SARS-CoV-2 cell entry and provides a possible pathway into the central nervous system. *bioRxiv* **2020**, 1–36.
 20. Galeotti, C.; Bayry, J. Autoimmune and inflammatory diseases following COVID-19. *Nat. Rev. Rheumatol.* **2020**, doi:10.1038/s41584-020-0448-7.
 21. Riphagen, S.; Gomez, X.; Gonzalez-Martinez, C.; Wilkinson, N.; Theocharis, P. Hyperinflammatory shock in children during COVID-19 pandemic. *Lancet (London, England)* **2020**, 6736, 2019–2020, doi:10.1016/S0140-6736(20)31094-1.
 22. Ace, H.; Prado, P.; Monteil, V.; Kwon, H.; Prado, P.; Hagelkru, A.; Wimmer, R.A.; Stahl, M.; Wirnsberger, G.; Zhang, H.; et al. Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell* **2020**, *181*, 905–913, doi:10.1016/j.cell.2020.04.004.
 23. Walls, A.C.; Park, Y.-J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Velesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, 1–12, doi:10.1016/j.cell.2020.02.058.
 24. Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T.S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, 1–10, doi:10.1016/j.cell.2020.02.052.
 25. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* **2020**, *3405*, 1–9, doi:10.1126/science.abb3405.
 26. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; et al. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* **2020**, doi:10.1038/s41586-020-2223-y.
 27. Wanchao, Y.; Chunyou, M.; Xiaodong, L.; Dan-Dan, S.; Qingya, S.; Haixia, S.; Xiaoxi, W.; Fulai, Z.; Wenfeng, Z.; Minqi, G.; et al. Structural Basis for the Inhibition of the RNA-Dependent RNA Polymerase from SARS- CoV-2 by Remdesivir. *Science* **2020**, *1560*, 1–30.
 28. Gordon, D.E.; Jang, G.M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K.M.; O'Meara, M.J.; Rezelj, V. V.; Guo, J.Z.; Swaney, D.L.; et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, 1–13, doi:10.1038/s41586-020-2286-9.
 29. Cao, B.; Wang, Y.; Wen, D.; Liu, W.; Wang, J.; Fan, G.; Ruan, L.; Song, B.; Cai, Y.; Wei, M.; et al. A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *N. Engl. J. Med.* **2020**, 1–13, doi:10.1056/NEJMoa2001282.

30. Zhang, X.; Tan, Y.; Ling, Y.; Lu, G.; Liu, F.; Yi, Z.; Jia, X.; Wu, M.; Wang, J.; Xu, M.; et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature* **2020**, doi:10.1038/s41586-020-2355-0.
31. Li, G.; De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat. Rev. Drug Discov.* **2020**, *19*, 149–150, doi:10.1038/d41573-020-00016-0.
32. Narayanan, K.; Huang, C.; Makino, S. SARS coronavirus accessory proteins. *Virus Res.* **2008**, *133*, 113–121, doi:10.1016/j.virusres.2007.10.009.
33. Liu, D.X.; Fung, T.S.; Chong, K.K.L.; Shukla, A.; Hilgenfeld, R. Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral Res.* **2014**, *109*, 97–109, doi:10.1016/j.antiviral.2014.06.013.
34. Yang, Y.; Zhang, L.; Geng, H.; Deng, Y.; Huang, B.; Guo, Y.; Zhao, Z.; Tan, W. The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. *Protein Cell* **2013**, *4*, 951–961, doi:10.1007/s13238-013-3096-8.
35. Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **2017**, *22*, 2–4, doi:10.2807/1560-7917.ES.2017.22.13.30494.
36. Velazquez-Salinas, L.; Zarate, S.; Eberl, S.; Gladue, D.P.; Novella, I.; Borca, M. V Positive selection of ORF3a and ORF8 genes drives the evolution of SARS-CoV-2 during the 2020 COVID-19 pandemic. *bioRxiv* **2020**, 2020.04.10.035964, doi:10.1101/2020.04.10.035964.
37. Oostra, M.; de Haan, C.A.M.; Rottier, P.J.M. The 29-Nucleotide Deletion Present in Human but Not in Animal Severe Acute Respiratory Syndrome Coronaviruses Disrupts the Functional Expression of Open Reading Frame 8. *J. Virol.* **2007**, *81*, 13876–13888, doi:10.1128/jvi.01631-07.
38. Yiwen Zhang, Junsong Zhang, Yingshi Chen, Baohong Luo, Yaocang Yuan, Feng Huang, Tao Yang, Fei Yu, Jun Liu, Bingfen Liu, Zheng Song, Jingliang Chen, Ting Pan, Xu Zhang, Yuzhuang Li, Rong Li, Wenjing Huang, Fei Xiao, and H.Z. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. *bioRxiv* **2020**.
39. Li, J.-Y.; Liao, C.-H.; Wang, Q.; Tan, Y.-J.; Luo, R.; Qiu, Y.; Ge, X.-Y. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. *Virus Res.* **2020**, 198074, doi:10.1016/j.virusres.2020.198074.
40. Ksiazek, T.G.; Erdman, D.; Goldsmith, C.S.; Zaki, S.R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J.A.; Lim, W.; et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **2003**, *348*, 1394–1398, doi:10.1056/NEJMoa030781.
41. Guan, Y.; Zheng, B.J.; He, Y.Q.; Liu, X.L.; Zhuang, Z.X.; Cheung, C.L.; Luo, S.W.; Li, P.H.; Zhang, L.J.; Guan, Y.J.; et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. *Science* **2003**, *302*, 276–278, doi:10.1126/science.1087139.
42. Consortium, T.C.S.M.E. Molecular Evolution of the SARS Coronavirus during the Course of the SARS Epidemic in China. *Science* **2004**, *303*, 1666–1669.
43. Koh, D.; Lim, M.; Ong, C.; Chia, S. Tracing SARS- Coronavirus Variant with Large Genomic Deletion. *Emerg. Infect. Dis.* **2005**, *11*, 168–170.
44. Muth, D.; Corman, V.M.; Roth, H.; Binger, T.; Dijkman, R.; Gottula, L.T.; Gloza-Rausch, F.; Balboni, A.; Battilani, M.; Rihtarič, D.; et al. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human

- transmission. *Sci. Rep.* **2018**, *8*, 1–11, doi:10.1038/s41598-018-33487-8.
45. LaRinda A. Holland, Emily A. Kaelin, Rabia Maqsood, Bereket Estifanos, Lily I. Wu, Arvind Varsani, Rolf U. Halden, Brenda G. Hogue, Matthew Scotch, E.S.L. An 81 nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (Jan-Mar 2020). *J. Virol.* **2020**, doi:10.1128/JVI.00711-20.
 46. Yvonne CF Su, Danielle E Anderson, Barnaby E Young, Feng Zhu, Martin Linster, Shirin Kalimuddin, Jenny GH Low, Zhuang Yan, Jayanthi Jayakumar, Louisa Sun, Gabriel Yan, Ian H Mendenhall, Yee-Sin Leo, David Chien Lye, Lin-Fa Wang, G.J.S. Discovery of a 382-nt deletion during the early evolution of SARS-CoV-2. *bioRxiv* **2020**.
 47. Gong, Y.-N.; Tsao, K.-C.; Hsiao, M.-J.; Huang, C.-G.; Huang, P.-N.; Huang, P.-W.; Lee, K.-M.; Liu, Y.-C.; Yang, S.-L.; Kuo, R.-L.; et al. SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. *Emerg. Microbes Infect.* **2020**, 1–37, doi:10.1080/22221751.2020.1782271.
 48. Worobey, M.; Holmes, E.C. Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **1999**, *80*, 2535–2543, doi:10.1099/0022-1317-80-10-2535.
 49. Hatcher, E.L.; Zhdanov, S.A.; Bao, Y.; Blinkova, O.; Nawrocki, E.P.; Ostapchuck, Y.; Schaffer, A.A.; Rodney Brister, J. Virus Variation Resource-improved response to emergent viral outbreaks. *Nucleic Acids Res.* **2017**, *45*, D482–D490, doi:10.1093/nar/gkw1065.
 50. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797, doi:10.1093/nar/gkh340.
 51. Chen, S.; Zheng, X.; Zhu, J.; Ding, R.; Jin, Y.; Zhang, W.; Yang, H.; Zheng, Y.; Li, X.; Duan, G. Extended ORF8 Gene Region Is Valuable in the Epidemiological Investigation of Severe Acute Respiratory Syndrome – Similar Coronavirus. *J. Infect. Dis.* **2020**, *222*, 223–233, doi:10.1093/infdis/jiaa278.
 52. Zwart, M.P.; Elena, S.F. Matters of Size : Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu. Rev. Virol.* **2015**, *2*, 161–179, doi:10.1146/annurev-virology-100114-055135.
 53. Mccrone, J.T.; Lauring, A.S. Genetic bottlenecks in intraspecies virus transmission. *Curr. Opin. Virol.* **2018**, *28*, 20–25, doi:10.1016/j.coviro.2017.10.008.
 54. Joseph, S.B.; Swanstrom, R.; Kashuba, A.D.M.; Cohen, M.S. Bottlenecks in HIV - 1 transmission : insights from the study of founder viruses. *Nat. Rev. Microbiol.* **2015**, *13*, 414–425, doi:10.1038/nrmicro3471.
 55. Tsetsarkin, K.A.; Chen, R.; Leal, G.; Forrester, N.; Higgs, S.; Huang, J.; Weaver, S.C. Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. *Proc. Natl. Acad. Sci.* **2011**, *108*, 7872–7877, doi:10.1073/pnas.1018344108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1018344108.
 56. Eal, L.E.A.R. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5095–5100.
 57. Hundertmark, K.J.; Daele, L.J. Van Founder effect and bottleneck signatures in an introduced , insular population of elk. *Conserv. Genet.* **2010**, *11*, 139–147, doi:10.1007/s10592-009-0013-z.
 58. Polymorphisms, H.I. V; Bhattacharya, T.; Daniels, M.; Heckerman, D.; Foley, B.; Frahm, N.; Kadie, C.; Carlson, J.; Yusim, K.; McMahon, B.; et al. Founder Effects in the Assessment of HIV Polymorphisms and HLAAllele Associations. *Science* **2007**, *315*, 1583–1587.
 59. Farkas, C.; Fuentes-villalobos, F.; Garrido, J.L.; Jody, J. Insights on early mutational events in

- SARS-CoV-2 virus reveal founder effects across geographical regions. *bioRxiv* **2020**.
60. Lefkowitz, E.J.; Dempsey, D.M.; Hendrickson, R.C.; Orton, R.J.; Siddell, S.G.; Smith, D.B. Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **2018**, *46*, D708–D717, doi:10.1093/nar/gkx932.
 61. Azhar, E.I.; El-Kafrawy, S.A.; Farraj, S.A.; Hassan, A.M.; Al-Saeed, M.S.; Hashem, A.M.; Madani, T.A. Evidence for camel-to-human transmission of MERS coronavirus. *N. Engl. J. Med.* **2014**, *370*, 2499–2505, doi:10.1056/NEJMoa1401505.
 62. Lau, S.K.P.; Feng, Y.; Chen, H.; Luk, H.K.H.; Yang, W.-H.; Li, K.S.M.; Zhang, Y.-Z.; Huang, Y.; Song, Z.-Z.; Chow, W.-N.; et al. Severe Acute Respiratory Syndrome (SARS) Coronavirus ORF8 Protein Is Acquired from SARS-Related Coronavirus from Greater Horseshoe Bats through Recombination. *J. Virol.* **2015**, *89*, 10532–10547, doi:10.1128/jvi.01048-15.
 63. Woo, P.C.Y.; Lau, S.K.P.; Lam, C.S.F.; Lau, C.C.Y.; Tsang, A.K.L.; Lau, J.H.N.; Bai, R.; Teng, J.L.L.; Tsang, C.C.C.; Wang, M.; et al. Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavi. *J. Virol.* **2012**, *86*, 3995–4008, doi:10.1128/jvi.06540-11.
 64. Latinne, A.; Hu, B.; Olival, K.J.; Zhu, G.; Zhang, L.; Li, H.; Chmura, A.A.; Field, H.E.; Zambrana-Torrel, C.; Epstein, J.H.; et al. Origin and cross-species transmission of bat coronaviruses in China. *bioRxiv* **2020**, 2020.05.31.116061, doi:10.1101/2020.05.31.116061.
 65. Banerjee, A.; Kulcsar, K.; Misra, V.; Frieman, M.; Mossman, K. Bats and coronaviruses. *Viruses* **2019**, *11*, 7–9, doi:10.3390/v11010041.
 66. Epstein, J.H.; McEachern, J.; Zhang, J.; Daszak, P.; Wang, H.; Field, H.; Li, W.; Eaton, B.T.; Wang, L.-F.; Yu, M.; et al. Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science* (80-.). **2005**, *310*, 676–679, doi:10.1109/NEMS.2006.334722.
 67. Wu, Z.; Yang, L.; Ren, X.; Zhang, J.; Yang, F.; Zhang, S.; Jin, Q. ORF8-related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J. Infect. Dis.* **2016**, *213*, 579–583, doi:10.1093/infdis/jiv476.
 68. Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezuk, Y.; McGinnis, S.; Madden, T.L. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **2008**, *36*, 5–9, doi:10.1093/nar/gkn201.
 69. Wrapp, D.; Wang, N.; Corbett, K.S.; Goldsmith, J.A.; Hsieh, C.L.; Abiona, O.; Graham, B.S.; McLellan, J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263, doi:10.1126/science.aax0902.
 70. Lau, S.K.P.; Woo, P.C.Y.; Li, K.S.M.; Huang, Y.; Tsoi, H.W.; Wong, B.H.L.; Wong, S.S.Y.; Leung, S.Y.; Chan, K.H.; Yuen, K.Y. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14040–14045, doi:10.1073/pnas.0506735102.
 71. Xiao, K.; Zhai, J.; Feng, Y.; Zhou, N.; Zhang, X.; Zou, J.-J.; Li, N.; Guo, Y.; Li, X.; Shen, X.; et al. Isolation and Characterization of 2019-nCoV-like Coronavirus from Malayan Pangolins. *bioRxiv* **2020**, 2020.02.17.951335, doi:10.1101/2020.02.17.951335.
 72. Zhang, T.; Wu, Q.; Zhang, Z. Pangolin homology associated with 2019-nCoV. *bioRxiv* **2020**, 2020.02.19.950253, doi:10.1101/2020.02.19.950253.
 73. Liu, P.; Jiang, J.Z.; Wan, X.F.; Hua, Y.; Li, L.; Zhou, J.; Wang, X.; Hou, F.; Chen, J.; Zou, J.; et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog.* **2020**, *16*, 1–13, doi:10.1371/journal.ppat.1008421.

74. Jiang, B.; Wei, W.; Yuan, T.; Zheng, K.; Cui, X.; Li, J.; Pei, G.; Qiang, X.; Cheung, W.Y.; Li, L.; et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **2020**, *583*, doi:10.1038/s41586-020-2169-0.
75. Andersen, K.G.; Rambaut, A.; Lipkin, W.I.; Holmes, E.C.; Garry, R.F. The proximal origin of SARS-CoV-2. *Nat. Med.* **2020**, *26*, 450–452, doi:10.1038/s41591-020-0820-9.
76. Li, X.; Giorgi, E.E.; Marichannegowda, M.H.; Foley, B.; Xiao, C.; Kong, X.-P.; Chen, Y.; Gnanakaran, S.; Korber, B.; Gao, F. Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci. Adv.* **2020**, eabb9153, doi:10.1126/sciadv.abb9153.
77. Lole, K.S.; Bollinger, R.C.; Paranjape, R.S.; Gadkari, D.; Kulkarni, S.S.; Novak, N.G.; Ingersoll, R.; Sheppard, H.W.; Ray, S.C. Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *J. Virol.* **1999**, *73*, 152–160, doi:10.1128/jvi.73.1.152-160.1999.
78. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **2015**, *1*, 1–5, doi:10.1093/ve/vev003.
79. Wong, M.C.; Cregeen, S.J.J.; Ajami, N.J.; Petrosino, J.F. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. *bioRxiv* **2020**, 2013, 2020.02.07.939207, doi:10.1101/2020.02.07.939207.
80. Juan Angel Patino-Galindo, Ioan Filip, Mohammed AlQuraishi, R.R. Recombination and convergent evolution led to the emergence of 2019 Wuhan coronavirus. *bioRxiv* **2020**, 1–13.
81. Xiang, D.; Shen, X.; Pu, Z.; Irwin, D.M.; Liao, M.; Shen, Y. Convergent evolution of human-isolated H7N9 avian influenza A viruses. *J. Infect. Dis.* **2018**, *217*, 1699–1707, doi:10.1093/infdis/jiy082.
82. Qiang Ni, S.M. and J.Z. Template-based and free modeling of I-TASSER and QUARK pipelines using predicted contact maps in CASP12. *Proteins* **2018**, *176*, 139–148, doi:10.1016/j.physbeh.2017.03.040.
83. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710, doi:10.1038/s41586-019-1923-7.
84. Heo, L.; Feig, M. Modeling of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Proteins by Machine Learning and Physics-Based Refinement. *bioRxiv* **2020**, 2, 2020.03.25.008904, doi:10.1101/2020.03.25.008904.
85. Sung, S.; Chao, C.; Jeng, K.; Yang, J.; Lai, M.M.C. The 8ab protein of SARS-CoV is a luminal ER membrane-associated protein and induces the activation of ATF6. *Virology* **2009**, *387*, 402–413, doi:10.1016/j.virol.2009.02.021.
86. Robinson, C. V.; Nagai, K.; Janda, C.Y.; Li, J.; Oubridge, C.; Herna, H. Recognition of a signal peptide by the signal recognition particle. *Nature* **2010**, *465*, doi:10.1038/nature08870.
87. Signal Peptide Recognition. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 2893, doi:10.1038/nrm2893.
88. Martoglio, B. Signal sequences : more than just greasy peptides. *Trends Cell Biol.* **1998**, *8*, 14119–14123.
89. Feige, M.J.; Hendershot, L.M. Disulfide bonds in ER protein folding and homeostasis. *Curr. Opin. Cell Biol.* **2011**, *23*, 167–175, doi:10.1016/j.ceb.2010.10.012.Disulfide.
90. Benham, A.M. Protein Secretion and the Endoplasmic Reticulum. *Cold Spring Harb. Perspect. Biol.* **2012**.
91. Young-Mi Go, D.P.J. Redox compartmentalization in eukaryotic cells. *Biochim. Biophys. Acta*

- 2008, 1780, 1273–1290, doi:10.1016/j.bbagen.2008.01.011.Redox.
92. Barlowe, C.K.; Miller, E.A. Secretory Protein Biogenesis and Traffic in the Early Secretory Pathway. *Cell Struct. Traffick.* **2013**, *193*, 383–410, doi:10.1534/genetics.112.142810.
 93. Ellgaard, L.; Sevier, C.S.; Bulleid, N.J. How Are Proteins Reduced in the Endoplasmic Reticulum? *Trends Biochem. Sci.* **2018**, *43*, 32–43, doi:10.1016/j.tibs.2017.10.006.
 94. Le, T.M.; Wong, H.H.; Tay, F.P.L.; Fang, S.; Keng, C.T.; Tan, Y.J.; Liu, D.X. Expression, post-translational modification and biochemical characterization of proteins encoded by subgenomic mRNA8 of the severe acute respiratory syndrome coronavirus. *FEBS J.* **2007**, *274*, 4211–4222, doi:10.1111/j.1742-4658.2007.05947.x.
 95. Vinh, P.; Lam, N.; Goldman, R.; Karagiannis, K.; Narsule, T.; Simonyan, V.; Soika, V.; Mazumder, R. Structure-based Comparative Analysis and Prediction of N-linked Glycosylation Sites in Evolutionarily Distant Eukaryotes. *Genomics. Proteomics Bioinformatics* **2013**, *11*, 96–104, doi:10.1016/j.gpb.2012.11.003.
 96. Xu, C.; Ng, D.T.W. Glycosylation-directed quality control of protein folding. *Nat. Rev. Mol. Cell Biol.* **2015**, doi:10.1038/nrm4073.
 97. Li, F.; Li, C.; Wang, M.; Webb, G.I.; Zhang, Y.; Whisstock, J.C.; Song, J. GlycoMine : a machine learning-based approach for predicting N- , C- and O-linked glycosylation in the human proteome. *Bioinformatics* **2015**, *31*, 1411–1419, doi:10.1093/bioinformatics/btu852.
 98. Caragea, C.; Sinapov, J.; Silvescu, A.; Dobbs, D.; Honavar, V. Glycosylation site prediction using ensembles of Support Vector Machine classifiers. *BMC Bioinformatics* **2007**, *13*, 1–13, doi:10.1186/1471-2105-8-438.
 99. Bagdonaite, I.; Wandall, H.H. Global aspects of viral glycosylation. *Glycobiology* **2018**, *28*, 443–467, doi:10.1093/glycob/cwy021.
 100. R. Gupta, E.J. and S.B. Prediction of N-glycosylation sites in human proteins Available online: <http://www.cbs.dtu.dk/services/NetNGlyc/>.
 101. Watanabe, Y.; Bowden, T.A.; Wilson, I.A.; Crispin, M. BBA - General Subjects Exploitation of glycosylation in enveloped virus pathobiology. *BBA - Gen. Subj.* **2019**, *1863*, 1480–1497, doi:10.1016/j.bbagen.2019.05.012.
 102. Watanabe, Y.; Allen, J.D.; Wrapp, D.; McLellan, J.S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *9983*, 1–9, doi:10.1126/science.abb9983.
 103. Chen, C.; Ping, Y.; Lee, H.; Chen, K.; Lee, Y.; Chan, Y.; Lien, T.; Jap, T.; Lin, C.; Kao, L.; et al. Open Reading Frame 8a of the Human Severe Acute Respiratory Syndrome Coronavirus Not Only Promotes Viral Replication but Also Induces Apoptosis. *J. Infect. Dis.* **2007**, *112*, 405–415, doi:10.1086/519166.
 104. Law, P.Y.P.; Liu, Y.M.; Geng, H.; Kwan, K.H.; Waye, M.M.Y.; Ho, Y.Y. Expression and functional characterization of the putative protein 8b of the severe acute respiratory syndrome-associated coronavirus. *FEBS Lett.* **2006**, *580*, 3643–3648, doi:10.1016/j.febslet.2006.05.051.
 105. Yount, B.; Roberts, R.S.; Sims, A.C.; Deming, D.; Frieman, M.B.; Sparks, J.; Denison, M.R.; Davis, N.; Baric, R.S. Severe Acute Respiratory Syndrome Coronavirus Group-Specific Open Reading Frames Encode Nonessential Functions for Replication in Cell Cultures and Mice. *J. Virol.* **2005**, *79*, 14909–14922, doi:10.1128/JVI.79.23.14909.
 106. Keng, C.T.; Åkerström, S.; Leung, C.S.W.; Poon, L.L.M.; Peiris, J.S.M.; Mirazimi, A.; Tan, Y.J. SARS coronavirus 8b reduces viral replication by down-regulating E via an ubiquitin-

- independent proteasome pathway. *Microbes Infect.* **2011**, *13*, 179–188, doi:10.1016/j.micinf.2010.10.017.
107. Keng, C.T.; Choi, Y.W.; Welkers, M.R.A.; Chan, D.Z.L.; Shen, S.; Gee Lim, S.; Hong, W.; Tan, Y.J. The human severe acute respiratory syndrome coronavirus (SARS-CoV) 8b protein is distinct from its counterpart in animal SARS-CoV and down-regulates the expression of the envelope protein in infected cells. *Virology* **2006**, *354*, 132–142, doi:10.1016/j.virol.2006.06.026.
 108. Totura, A.L.; Baric, R.S. SARS coronavirus pathogenesis: host innate immune responses and viral antagonism of interferon. *Curr. Opin. Virol.* **2012**, *2*, 264–275.
 109. Hui, H.; Sing, T.; Fang, S.; Huang, M.; Tra, M.; Xiang, D. Accessory proteins 8b and 8ab of severe acute respiratory syndrome coronavirus suppress the interferon signaling pathway by mediating ubiquitin-dependent rapid degradation of interferon regulatory factor 3. *Virology* **2018**.
 110. Luo, H. Interplay between the virus and the ubiquitin – proteasome system : molecular mechanism of viral pathogenesis. *Curr. Opin. Virol.* **2015**, *17*, 1–10, doi:10.1016/j.coviro.2015.09.005.
 111. Moshe, A.; Gorovits, R. Virus-Induced Aggregates in Infected Cells. *Viruses* **2012**, *4*, 2218–2232, doi:10.3390/v4102218.
 112. Shi, C.S.; Nabar, N.R.; Huang, N.N.; Kehrl, J.H. SARS-Coronavirus Open Reading Frame-8b triggers intracellular stress pathways and activates NLRP3 inflammasomes. *Cell Death Discov.* **2019**, *5*, doi:10.1038/s41420-019-0181-7.
 113. Yang, Y.; Wang, H.; Kouadir, M.; Song, H.; Shi, F. Recent advances in the mechanisms of NLRP3 in inflammasome activation and its inhibitors. *Cell Death Dis.* **2019**, doi:10.1038/s41419-019-1413-8.
 114. Kanneganti, T. Central roles of NLRs and inflammasomes in viral infection. *Nat. Rev. Immunol.* **2010**, *10*, 688–698, doi:10.1038/nri2851.
 115. He, B. Viruses , endoplasmic reticulum stress , and interferon responses. *Cell Death Differ.* **2006**, *13*, 393–403, doi:10.1038/sj.cdd.4401833.
 116. Park, M.D. Immune evasion via SARS-CoV-2 ORF8 protein? *Nat. Rev. Immunol.* **2020**, *20*, 111823, doi:10.1038/s41577-020-0360-z.
 117. Jérôme Hadjadj Nader Yatim , Laura Barnabei, Aurélien Corneau, Jeremy Boussier, Nikaïa Smith, Hélène Péré, Bruno Charbit, Vincent Bondet, Camille Chenevier-Gobeaux, Paul Breillat , Nicolas Carlier, Rémy Gauzit, Caroline Morbieu, Frédéric Pène , Nathalie M, B.T. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **2020**, *6027*.
 118. Ma, J.; Zhang, X.; Soloveva, V.; Warren, T.; Guo, F.; Wu, S.; Lu, H.; Guo, J.; Su, Q.; Shen, H.; et al. Enhancing the antiviral potency of ER α -glucosidase inhibitor IHVR-19029 against hemorrhagic fever viruses in vitro and in vivo. *Antiviral Res.* **2018**, *150*, 112–122, doi:10.1016/j.antiviral.2017.12.008.Enhancing.
 119. Chang, J.; Warren, T.K.; Zhao, X.; Gill, T.; Guo, F.; Wang, L.; Ann, M.; Du, Y.; Alonzi, D.S.; Yu, W.; et al. Small molecule inhibitors of ER α -glucosidases are active against multiple hemorrhagic fever viruses. *Antiviral Res.* **2013**, *98*, 432–440, doi:10.1016/j.antiviral.2013.03.023.
 120. Li, J.; You, Z.; Wang, Q.; Zhou, Z.; Qiu, Y.; Luo, R.; Ge, X. The epidemic of 2019-novel-coronavirus (2019-nCoV) pneumonia and insights for emerging infectious diseases in the future. *Microbes Infect.* **2020**.

121. Hwang, J.; Qi, L. Quality Control in the Endoplasmic Reticulum : Crosstalk between ERAD and UPR pathways. *Trends Biochem. Sci.* **2018**, *43*, 593–605, doi:10.1016/j.tibs.2018.06.005.
122. Bernasconi, R.; Molinari, M. ERAD and ERAD tuning : disposal of cargo and of ERAD regulators from the mammalian ER. *Curr. Opin. Cell Biol.* **2011**, *23*, 176–183, doi:10.1016/j.ceb.2010.10.002.