

An issue of concern: unique truncated ORF8 protein variants of SARS-CoV-2

Sk. Sarif Hassan¹, Vaishnavi Kodakandla², Elrashdy M. Redwan³, Kenneth Lundstrom⁴, Pabitra Pal Choudhury⁵, Tarek Mohamed Abd El-Aziz⁶, Kazuo Takayama⁷, Ramesh Kandimalla⁸, Amos Lal⁹, Ángel Serrano-Aroca¹⁰, Gajendra Kumar Azad¹¹, Alaa A.A. Aljabali¹², Giorgio Palù¹³, Gaurav Chauhan¹⁴, Parise Adadi¹⁵, Murtaza Tambuwala¹⁶, Adam M. Brufsky¹⁷, Wagner Baetas-da-Cruz¹⁸, Debmalya Barh¹⁹, Vasco Azevedo²⁰, Nikolas G. Bazan²¹, Bruno Silva Andrade²², Raner José Santana Silva²³ and Vladimir N. Uversky²⁴

¹ Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram, India

² Department of Life sciences, Sophia College For Women, University of Mumbai, Mumbai, India

³ Faculty of Science, Department of Biological Science, King Abdulaziz University, Jeddah, Saudi Arabia

⁴ PanTherapeutics, Lutry, Switzerland

⁵ Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

⁶ Department of Cellular and Integrative Physiology, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

⁷ Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan

⁸ Applied Biology, CSIR-Indian Institute of Chemical Technology, Hyderabad, India

⁹ Division of Pulmonary and Critical Care Medicine, Mayo Clinic Rochester, Rochester, NY, United States

¹⁰ Biomaterials and Bioengineering Lab, Centro de Investigacion Traslacional San Alberto Magno, Universidad Catolica de Valencia San Vicente Martir, Valencia, Spain

¹¹ Department of Zoology, Patna University, Patna, India

¹² Department of Pharmaceutics and Pharmaceutical, Yarmouk University, Irbid, Jordan

¹³ Department of Molecular Medicine, University of Padova, Padova, Italy

¹⁴ School of Engineering and Sciences, Tecnológico de Monterrey, Monterrey, Mexico

¹⁵ Department of Food Science, University of Otago, University of Otago, Dunedin, New Zealand

¹⁶ School of Pharmacy and Pharmaceutical Science, Ulster University, Coleraine, UK

¹⁷ Department of Medicine, Division of Hematology/Oncology, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

¹⁸ Translational Laboratory in Molecular Physiology, Centre for Experimental Surgery, College of Medicine, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

¹⁹ Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and 46 Applied Biotechnology (IIOAB), Nonakuri, India

²⁰ Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

²¹ Neuroscience Center of Excellence, School of Medicine, LSU Health New Orleans, New Orleans, LA, United States

²² Laboratório de Bioinformática e Química Computacional, Departamento de Ciências Biológicas, Universidade Estadual do Sudoeste da Bahia, Jequié, Brazil

²³ Departamento de Ciências Biológicas (DCB), Programa de Pós-Graduação em Genética e Biologia Molecular (PPGGBM), Universidade Estadual de Santa Cruz (UESC), Ilheus, Brazil

²⁴ Department of Molecular Medicine, University of South Florida, Tampa, FL, United States

Submitted 16 September 2021

Accepted 27 February 2022

Published 21 March 2022

Corresponding author

Vladimir N. Uversky,

vuversky@usf.edu

Academic editor

Jun Chen

Additional Information and
Declarations can be found on
page 19

DOI [10.7717/peerj.13136](https://doi.org/10.7717/peerj.13136)

© Copyright

2022 Hassan et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

ABSTRACT

Open reading frame 8 (ORF8) shows one of the highest levels of variability among accessory proteins in Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of Coronavirus Disease 2019 (COVID-19). It was

previously reported that the ORF8 protein inhibits the presentation of viral antigens by the major histocompatibility complex class I (MHC-I), which interacts with host factors involved in pulmonary inflammation. The ORF8 protein assists SARS-CoV-2 in evading immunity and plays a role in SARS-CoV-2 replication. Among many contributing mutations, Q27STOP, a mutation in the ORF8 protein, defines the B.1.1.7 lineage of SARS-CoV-2, engendering the second wave of COVID-19. In the present study, 47 unique truncated ORF8 proteins (T-ORF8) with the Q27STOP mutations were identified among 49,055 available B.1.1.7 SARS-CoV-2 sequences. The results show that only one of the 47 T-ORF8 variants spread to over 57 geo-locations in North America, and other continents, which include Africa, Asia, Europe and South America. Based on various quantitative features, such as amino acid homology, polar/non-polar sequence homology, Shannon entropy conservation, and other physicochemical properties of all specific 47 T-ORF8 protein variants, nine possible T-ORF8 unique variants were defined. The question as to whether T-ORF8 variants function similarly to the wild type ORF8 is yet to be investigated. A positive response to the question could exacerbate future COVID-19 waves, necessitating severe containment measures.

Subjects Bioinformatics, Computational Biology

Keywords ORF8, SARS-CoV-2, COVID-19, Truncated, Intrinsically disordered region, Truncation mutation, Continent distribution

INTRODUCTION

The world is proceeding through a very difficult time due to the Coronavirus Disease 2019 (COVID-19), of which the causative agent is the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) ([Hu et al., 2021](#); [Yuen et al., 2020](#); [Matheson & Lehner, 2020](#); [Wu et al., 2020](#); [Fontanet et al., 2021](#)). There are nine open reading frames (ORFs), which encode accessory proteins important for the modulation of metabolism in infected host cells and innate immunity evasion *via* a complicated signalome and an interactome ([Ren et al., 2020](#); [Hassan et al., 2021c, 2020b](#); [Díaz, 2020](#); [Stukalov et al., 2021](#)). The ORF8 protein is one of the most rapidly evolving accessory proteins among the beta coronaviruses, not only due to its ability to interfere with host immune response, but also with regards to its several missense mutations detected to date ([Li et al., 2020](#); [Zinzula, 2021](#); [Hassan et al., 2021a](#); [Flower et al., 2021](#)). ORF8 directly interacts with major histocompatibility complex class I (MHC-I) both *in vitro* and *in vivo*, and is down-regulated, which impairs its ability to carry out antigen presentation and rendering infected cells less sensitive to lysis by cytotoxic T lymphocytes ([Zhang et al., 2021](#)). ORF8 suppresses type I interferon antiviral responses and interacts with host factors involved in pulmonary inflammation and fibrogenesis ([Zhang et al., 2021](#); [Rashid et al., 2021](#)). From all viral proteomes interacting with human metalloproteome, the ORF8 proteins interact with 10 out of 58 human metalloproteins ([Chasapis et al., 2021](#)). Both SARS-CoV-2 and SARS-CoV ORF8 proteins play crucial roles in virus pathophysiological events, and dysregulate the TGF- β pathway, which is involved in the tissue fibrosis

(Pereira, 2020). The functional implications of SARS-CoV-2 ORF8 had already gained immense attention and ORF8 is considered an important component of the immune evasion machinery (Li et al., 2020; Pereira, 2020; Mohammad et al., 2020; Alkhansa, Lakkis & El Zein, 2021). The SARS-CoV-2 ORF8 protein has less than twenty percent amino acid sequence homology with the SARS-CoV ORF8 protein, and represents a rapidly evolving viral protein (Flower et al., 2021; Velazquez-Salinas et al., 2020). A molecular framework for understanding the rapid evolution of ORF8, its contributions to COVID-19 pathogenesis, and the potential for its neutralization by antibodies have been supported by the structural analysis of the ORF8 protein (Hachim et al., 2020; Wang et al., 2020a). The crosstalk between SARS-CoV-2 or SARS-CoV infection and the host cell proteome at different levels may enable the identification of distinct and common molecular mechanisms (Zhang et al., 2021). Of note, SARS-CoV-2 ORF8 interacts with a significant number of host proteins related to the endoplasmic reticulum quality control, glycosylation, and extracellular matrix organization, although the mechanism of action of ORF8 concerning those interacting proteins is uncertain (Wang et al., 2020a; Wu et al., 2021).

The clade S, a subtype of SARS-CoV-2, was identified to possess the mutation L84S in the ORF8 protein sequence (Koyama, Platt & Parida, 2020; Mercatelli & Giorgi, 2020; Sengupta, Hassan & Choudhury, 2021). Presently, among many variants of SARS-CoV-2, the lineage B.1.1.7 carries a larger than usual number of genetic changes (Galloway et al., 2021; Ramírez et al., 2021; Frampton et al., 2021). Among many non-synonymous mutations, Q27STOP in the ORF8 protein contributed to the extrapolation of the branch leading to lineage B.1.1.7 (Perchetti et al., 2021; Li et al., 2021). The Q27STOP mutation inactivates the ORF8 protein favoring further downstream mutations, and could be responsible for the increased transmissibility of the B.1.1.7 variant (Galloway et al., 2021; Borges et al., 2021). The B.1.1.7 variant, being more transmissible than the wild-type SARS-CoV-2, was first detected in September 2020 in the UK (Shen et al., 2021; Walensky, Walke & Fauci, 2021). It began to spread rapidly by mid-December and was correlated with a significant increase in the number of SARS-CoV-2 infections in the UK and worldwide.

Functional implications on the immune surveillance due to the ORF8 truncation at position 27 remain unclear (Pereira, 2020). Therefore, it is of utmost importance to gain insight into the functionality of the truncated ORF8 protein variants to comprehend the B.1.1.7 lineage through theoretical and experimental characterization and genomic surveillance worldwide (Davies et al., 2021). The present study aimed at characterizing the unique variations of truncated ORF8 proteins (T-ORF8) due to the Q27STOP mutation. Further, this investigation differentiates a single T-ORF8 variant among 47 distinct unique T-ORF8 variants present in SARS-CoV-2 worldwide, as of May 20th, 2021. Several clusters of the unique T-ORF8 have been identified based on the various bioinformatics features and phylogenetic relationships, along with the emerging variants of the unique T-ORF8.

DATA ACQUISITION AND METHODS

A total of 49,055 truncated ORF8 protein (T-ORF8) sequences (complete) from five continents (Asia, Africa, Europe, South America, and North America) were downloaded in FASTA format from the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/>). Note that no T-ORF8 protein sequence was found from Oceania. Next, FASTA files were processed in MATLAB 2021a for extracting unique T-ORF8 protein sequences for each continent. Note that only 47 unique T-ORF8 protein sequences were found. Here “unique T-ORF8” refers to a T-ORF8 protein sequence, which is different from other T-ORF8 sequences by the arrangement of amino acids, showing a non-zero Hamming distance from other T-ORF8 sequences. Consequently, the amino acid sequence of a “unique T-ORF8” variant is non-identical to other T-ORF8 sequences.

Derivation of polar/non-polar sequences and associated phylogeny

Every amino acid in a given T-ORF8 sequence was identified as polar (Q) or non-polar (P). Thus, every unique T-ORF8 became a binary sequence with two symbols P and Q. Then, the homology of these sequences was determined using the Clustal Omega web-suite (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and then associated with the nearest neighbor phylogenetic relationship among the unique T-ORF8 variants. Further, unique T-ORF8 variants with distinct binary polar/non-polar sequences were extracted (*Hassan et al., 2020a; Broome & Hecht, 2000*).

Frequency distribution of amino acids and phylogeny

The frequency of each amino acid present in a T-ORF8 sequence was determined using standard bioinformatics routine in *Matlab-2021a*. For each T-ORF8 protein, a twenty-dimensional frequency-vector considering the frequency of standard twenty amino acids can be obtained. Based on this frequency distribution of amino acids several consequences were drawn. The distance (Euclidean metric) between any two pairs of frequency vectors was calculated for each pair of T-ORF8 sequences. The distance matrix was used to develop a phylogenetic relationship based on the nearest neighbor-joining method using the standard routine in *Matlab-2021a* (*Hassan et al., 2020c; Hassan, Choudhury & Roy, 2021*).

Deep phylogenetic analyses

All phylogenetic data preparation and calculations were conducted in MEGA X (*Kumar et al., 2018; Stecher, Tamura & Kumar, 2020*). For deep phylogeny analyses we used two nucleotide datasets: one composed by the alignment of all 47 truncated ORF 8 sequences in addition to the RATG13 ORF8 as an external group, and a second group using all 47 truncated ORF 8, the RATG13 ORF 8 as well as 66 representative ORF 8 sequences for different Alpha, Beta, Gamma, Delta, Mu, GH/490R and Omicron variants. The phylogeny estimations were inferred by Maximum Likelihood using the Hasegawa-Kishino-Yano model (*Hasegawa, Kishino & Yano, 1985*). A discrete Gamma distribution was used compute the evolutionary rate differences among sites: four categories +G,

parameter = 200.0000 and 500 bootstrap replications (*Benvenuto et al., 2020*). Furthermore, initial trees were heuristically obtained using Neighbor-Join and BioNJ algorithms with a Maximum Composite Likelihood (MCL) approach.

Amino acid conservation in terms of Shannon's entropy

The degree of conservation of amino acids embedded in a T-ORF8 protein was obtained by the well-known information-theoretic measure called “*Shannon's entropy* (SE)”. For each T-ORF8 protein, Shannon's entropy of amino acid conservation over the amino acid sequence of T-ORF8 protein was calculated using as described in (*Hassan et al., 2020c, 2021b*). For a given T-ORF8 sequence of length l (here $l = 26$), the conservation of amino acids was calculated as follows:

$$SE = - \sum_{i=1}^{20} p_{s_i} \log_{20}(p_{s_i})$$

where $p_{s_i} = \frac{k_i}{l}$; k_i represents the number of occurrences of an amino acid s_i in the T-ORF8 sequence (*Strait & Dewey, 1996*).

Prediction of molecular and physicochemical properties

Theoretical pI (PI), extinction coefficient (EC), instability index (II), aliphatic index (AI), protein solubility (PS), grand average of hydropathicity (GRAVY), and the number of tiny, small, aliphatic, aromatic, non- polar, polar, charged, basic and acidic residues of all unique T-ORF8 proteins were calculated using the web-servers 'ProtParam', 'Protein-sol' and EMBOSS Pepstats (*Gasteiger et al., 2005; Hebditch et al., 2017; Madeira et al., 2019*).

Intrinsic disorder analysis

All 47 T-ORF8 variants were subjected to the per-residue disorder analysis, for which PONDR[®] VSL2 algorithm was employed (*Obrodovic et al., 2005*). This tool showed good performance on proteins containing both structure and disorder and was favorably ranked in a recent Critical Assessment of protein Intrinsic Disorder prediction (CAID) experiment (*Necci, Piovesan & Tosatto, 2021*).

Finding functional motifs

The Eukaryotic Linear Motif (ELM) resource (<http://elm.eu.org/>) was used for finding functional sites in proteins (*Kumar et al., 2020*). ELMs (also known as short linear motifs (SLiMs)), are short protein interaction sites, which are commonly found in intrinsically disordered regions of proteins and define a wide range of protein functionality.

RESULTS

All 49,055 unique T-ORF8 protein variants were segregated from a set of available truncated ORF8 protein sequences collected from the NCBI database. Further, variability and commonality of the unique T-ORF8 proteins were analyzed from various quantitative measures as discussed in the methods section.

Table 1 Frequency and percentages unique T-ORF8 variants (continent-wise).*Percentages of the unique T-ORF8 variants on continents*

Continent	Total T-ORF8 (T)	Unique T-ORF8 (U)	Percentage, continent-wise	Percentage, world-wide
<i>Africa</i>	108	1	0.926%	1.96%
<i>Asia</i>	99	1	1.01%	1.96%
<i>Europe</i>	156	1	0.641%	1.96%
<i>South America</i>	1	1	100%	1.96%
<i>North America</i>	48,691	47	0.096%	92.16%
<i>World-wide</i>	49,055	47	0.104%	

Note:

Here 'U' stands for the total number of unique T-ORF8 variants over the total available T-ORF8 sequences, which is denoted by 'T'.

Table 2 Truncated ORF8 variants of length other than 26 amino acids.

Accession ID	Length (number of amino acid residues)	Date of collection	Geo-location	Remarks
QXX22250.1	22	20-10-2020	USA: KS	2*Identical sequence
QXX22346.1	22	24-09-2020	USA: MO	
QVF74147.1	24	27-04-2021	USA: Colorado	Worldwide frequency: 01
QRE01295.1	40	13-12-2020	USA: MD	Worldwide frequency: 01
QXX21038.1	41	30-10-2020	USA: OK	Worldwide frequency: 01
QLJ58176.1	41	09-04-2020	USA	2*Identical sequence
QLJ58236.1	41	16-04-2020	USA	

Characteristics of the unique variants of T-ORF8

The number of total sequences for each continent, the unique truncated ORF8 (T-ORF8) sequences, and their percentages are presented in [Table 1](#). The results showed that 47 unique T-ORF8 proteins among the total of 48,691 were present in North America ([Table 1](#)). The unique T-ORF8 variants from Africa, Asia, Europe, and South America were contained in the set of unique T-ORF8 variants available in North America. Additionally, there were seven T-ORF8 with amino acid lengths 22, 24, 40 and 41 available in North America as of May 18, 2021 ([Table 2](#)). Note that among the seven T-ORF8 sequences, only five were found to be unique, as mentioned in [Table 2](#). As of May 18, 2021 a single copy of the T-ORF8 proteins of amino acid lengths of 24 and 41 ([Table 2](#)) were found. There were two T-ORF8 variants of 41 amino acids available in North America. The most frequent T-ORF8 proteins so far observed, were the T-ORF8 proteins of 26 amino acids. It was observed that the T-ORF8 arose due to truncation at the residue positions 23, 25, 27, 41, and 42 of the 121 amino acid full-length ORF8 protein. We investigated the possible mutations for such truncations. A snapshot of the amino acid residues and their possible mutations with respect to the reference sequence NC 045512 is presented in [Fig. S1](#).

Note that at four positions, 23, 25, 27, and 40, amino acids glutamine (Gln) and cysteine (Cys) were truncated due to mutations at the codon's first and third positions, respectively. The amino acid valine (Val) was truncated due to three mutations at the third, second, and first positions of the codon 'GUG'. Furthermore, it was observed that the mutations at the positions 23 and 25 were identical (C to U), and the changes of bases were transition mutations; *i.e.*, pyrimidine (purine) to pyrimidine (purine). In contrast, the changes of bases of the truncated mutations at the positions 25 and 41 were transversal mutations; *i.e.*, pyrimidine (purine) to purine (pyrimidine). For position 42, three sequences of mutations were hypothesized, taking place at first, second, and third positions of the codon (GUG); *i.e.*, transition mutations (purine to purine), transversal mutation (pyrimidine to purine), and transversal mutation (purine to pyrimidine), respectively.

The list of unique T-ORF8 sequences of 26 amino acids with their representative accession IDs and sequences is presented in [Table S1](#). Further, it was found that the unique T-ORF8 variants from Africa, Asia, Europe and South America were identical in relation to P15, as illustrated in [Table S1](#). The date of sample collection, geo-location and accession ID of the first identified SARS-CoV-2 containing unique T-ORF8 variants are presented in [Table S2](#).

Based on [Table S2](#), the ORF8 protein sequence P15 was found in 48395 copies of the B.1.17 SARS-CoV-2 lineage in North America. Besides, the P15 variant with the Q27STOP mutation in the B.1.1.7 lineage was found in Africa, Asia, Europe and South America with a frequency of 108, 99, 156, and 1, respectively. None of the other 46 T-ORF8 unique variants was found on any continent, as of May 18, 2021. So, 46 unique T-ORF8 sequences were exclusively found in North America. Therefore, the P15 TORF8 variant is of particular interest for its uniqueness due to its apparent prevalence in most of the B.1.17 lineages of SARS-CoV-2 from North America and other continents.

In Europe, the P15 variant was first detected in two infected patients from Poland on March 15, 2020. In North America, two patients from Maryland were infected with the same SARS-CoV-2 P15 variant on May 27, 2020. After five days of the second occurrence of P15 in North America, one patient from Punjab-Pakistan (Asia) was infected by the P15 SARS-CoV-2 variant. Six months thereafter, the same variant was found in a patient from Ghana, for the first time in Africa. Twenty days after the fifth occurrence in Africa, the P15 variant was identified in Peru (South America) on December 31, 2020.

Additionally, the frequency distribution of the T-ORF8 P15 variants across the North American continent is presented in [Table 3](#). It was found that the T-ORF8 P15 variant spread over three geo-locations, Michigan, Florida, and Minnesota, with the highest number of frequencies of 5084, 6884, and 7416, respectively. The P15 variant was found for the first time in Maryland, but the frequency at this geo-location was 1171 on May 14, 2021.

There were 18 geo-locations, where the frequency of spread of the P15 variant was found to be less than 100 ([Table 3](#)). Among other geo-locations, Guam (US territory located in the Pacific Ocean) and North Dakota, USA, had the least number of patients infected by the B.1.1.7 variant containing the P15 protein. In Guam, according to the NCBI SARS-CoV-2 database, all seven patients were infected by the B.1.1.7 variant containing

Table 3 Distribution of cumulative frequency of P15 variants across North America.

Geo-location	Frequency	Geo-location	Frequency	Geo-location	Frequency
Wyoming	56	North Carolina	776	Iowa	141
Wisconsin	383	New York	887	Indiana	823
West Virginia	289	New Mexico	250	Illinois	1,426
Washington	83	New Jersey	1,815	Idaho	85
Virginia	917	New Hampshire	234	Hawaii	16
Vermont	209	Nevada	157	Guam	7
Utah	97	Nebraska	105	Georgia	1,232
Texas	3,420	Montana	25	Florida	6,884
Tennessee	993	Missouri	254	District of Columbia	61
South Dakota	86	Mississippi	40	Delaware	70
South Carolina	261	Minnesota	7,416	Connecticut	496
Rhode Island	339	Michigan	5,084	Colorado	533
Puerto Rico	224	Massachusetts	2,761	California	1,727
Pennsylvania	3,285	Maryland	1,171	CA, Santa Clara County	4
Oregon	166	Maine	79	CA, Humboldt	20
Oklahama	81	Louisiana	223	Arkansas	62
Ohio	1,191	Kentucky	145	Arizona	290
North Dakota	13	Kansas	100	Alaska	65
				Alabama	168
				USA	654

Table 4 Frequency distribution of unique T-ORF8 variants over the USA.

USA: states	Unique T-ORF8 variants	USA: states	Unique T-ORF8 variants
USA: California	P1, P30, P40	USA: Missouri	P28
USA: Connecticut	P32, P33	USA: New Jersey	P10, P41
USA: Florida	P4, P14, P16	USA: Ohio	P2
USA: Georgia	P21	USA: North Carolina	P47
USA: Illinois	P18	USA: Pennsylvania	P7, P17, P19, P20, P23, P27, P39
USA: Kentucky	P36	USA: Puerto Rico	P24
USA: Louisiana	P31	USA: Tennessee	P5, P22
USA: Maryland	P15, P26, P29, P34	USA: Rhode Island	P35
USA: Massachusetts	P35, P42	USA: Texas	P6, P9, P11, P25
USA: Michigan	P8, P38, P43, P46	USA: Utah	P45
USA: Minnesota	P3, P12, P13, P37, P44		

the P15, within a short period from February 21 to April 11, 2021. Also, in North Dakota, 13 of 16 patients were infected by the same strain of SARS-CoV-2 from February 2, 2021 to April 28, 2021.

The frequency distribution of all T-ORF8 variants across the US is presented in [Table 4](#). It is evident that all 47 unique T-ORF8 variants were detected in 21 different states of the

US. The highest number (7) of the unique T-ORF8 variants was detected as a first instance in Pennsylvania within a short period (March 2 to April 20, 2021) (Table 4). The P35 variant was found initially in two states: Rhode Island and Massachusetts on March 20, 2021. Furthermore, it was observed that all T-ORF8 variants other than P15 emerged for the first time from February 12, 2021 to April 28, 2021.

Using the Clustal Omega web-server, an amino acid sequence-based alignment and corresponding phylogenetic tree of the unique T-ORF8 variants are presented in Fig. S2. From the sequence alignment, it was derived that all unique T-ORF8 variants share identical amino acids methionine (Met), lysine (Lys), Gln, serine (Ser), and leucine (Leu) at positions 1, 2, 18, 21, and 22 respectively. Further it was found that T-ORF8 P15 is closest to the ORF8 sequences P13 and P14 (Fig. S2). Note that P15 was placed at the leftmost branch of the phylogenetic tree, which made the sequence P15 distinguishable from the rest of the T-ORF8 variants. The pairs of T-ORF8 variants (P13, P14), (P5, P6), (P33, P45), (P9, P37), (P19, P20), (P35, P36), (P31, P34), (P29, P43), (P27, P42), (P25, P26), (P22, P23), (P21, P40), (P17, P18), (P2, P8), and (P1, P47) were found to be the closest to each other based on the amino acid sequence homology-based phylogeny (Fig. S2).

Next, we conducted deep phylogenetic analysis. The first nucleotide dataset used for this analysis included the alignment of all 47 truncated ORF8 sequences in addition to the RATG13 ORF8 as an external group. Phylogenetic analysis of this group returned a tree (Fig. 1) with the highest log likelihood of -400.62 in an analysis with 48 nucleotide sequences and a total of 83 positions in the final dataset. In addition, Fig. 1 shows nine well-defined monophyletic groups, where the ORF 8 sequences 15 and 40 formed a unique clade in the group II. Based on this tree analysis it can be inferred that at least nine truncated ORF8 variants could be related to a specific SARS-CoV-2 variant and/or a specific variant subtype, in this case more related to the Alpha type. Interestingly, sequences 35 and 46 rooted outside the tree together with the external group ORF8 from the RATG13 genome. In these cases, the sequences 35 and 46 are possibly mutating to the corresponding SARS-CoV-2 start point or they could be related to a subtype with a small viral load and spread.

In the second dataset there were all 47 truncated ORF8 variants, the RATG13 ORF8, as well as 66 representative ORF8 sequences from different Alpha, Beta, Gamma, Delta, Mu, GH/490R, and Omicron variants. Results of the phylogenetic analysis of this set containing all 47 sequences and different SARS-CoV-2 ORF8 variant representatives are shown in Fig. S3, where the resultant tree presented a highest log likelihood of -430.29 .

This analysis involved 114 nucleotide sequences with a total of 81 positions in the final dataset. In this tree, one can find at least 14 group clades, and, as expected, almost all ORF8 variants are grouped together in the group I. The other 47 truncated ORF8 sequences presented similar tree topologies in comparison to the first dataset tree (cf. Figs. 1 and S3). On the other hand, the group IV showed that the sequences 15 and 40 formed a clear clade to the Alpha ORF8, what was expected. Furthermore, these sequences grouped with the ORF8 variant B.1.640, which was described as the new variant IHU found in France in December 2021 (Colson et al., 2021).

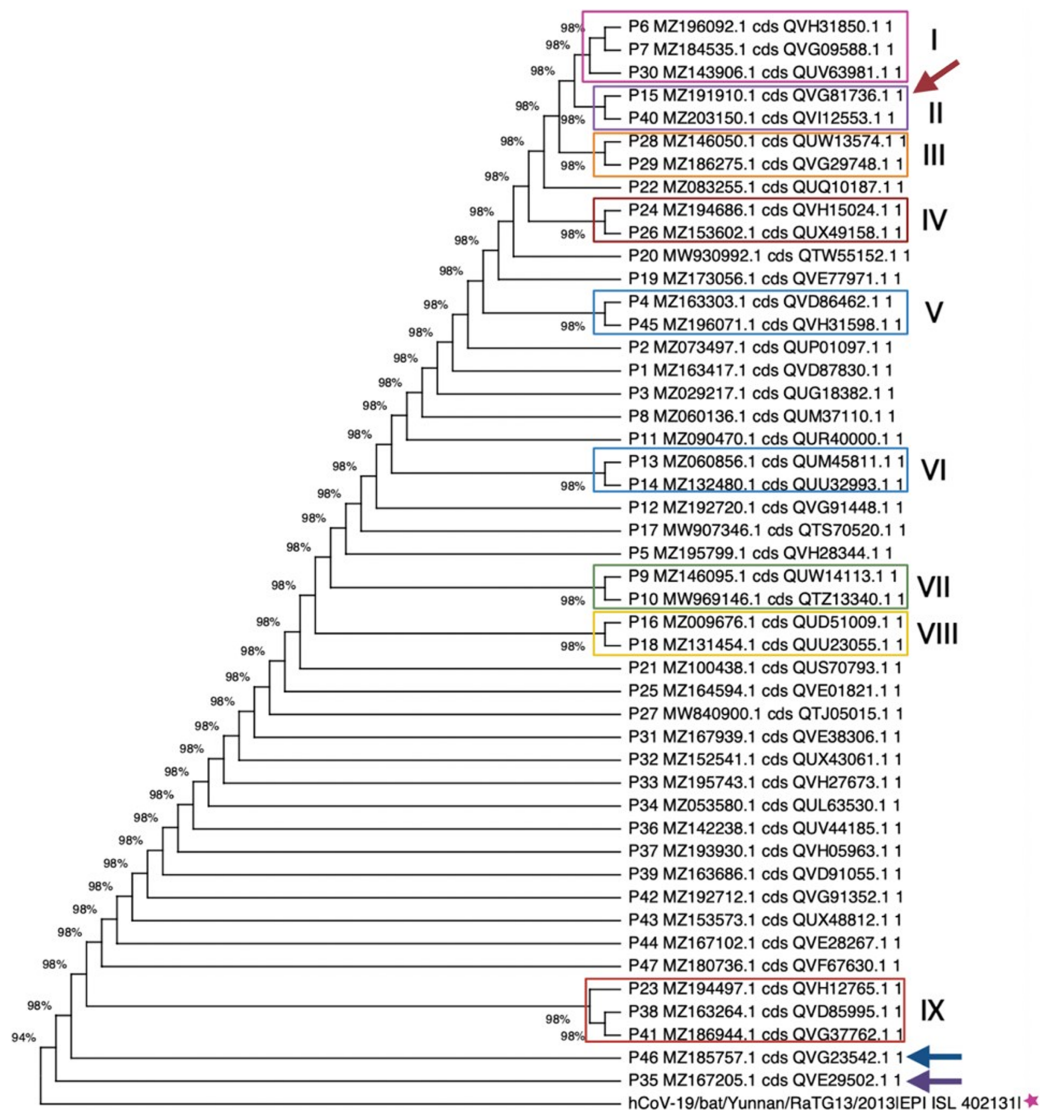


Figure 1 Maximum likelihood phylogenetic tree for the 47 truncated ORF8, using 500 bootstrap replications and the Hasegawa-Kishino-Yano model. Nine group clades were found, while sequences 35 and 46 (marine blue and purple arrows, respectively) are phylogenetically near to the RATG13 ORF8 sequence. Sequence 15 is indicated by a red arrow. [Full-size !\[\]\(1663bb69f307a960345edb0e712f8c02_img.jpg\) DOI: 10.7717/peerj.13136/fig-1](https://doi.org/10.7717/peerj.13136/fig-1)

Evaluation of intrinsic disorder content of 47 T-ORF8 proteins

Unfortunately, no structural information is available for the truncated forms of ORF8. In fact, although two X-ray structures were reported for the dimeric form of the SARS-CoV-2 ORF8 protein, both of these structures were solved for the ORF8 sequences containing residues 18–121; *i.e.*, they do not include information for more than a half (63%) of the N-terminal region of the full-length ORF8, which constitutes T-ORF8. Furthermore, truncated forms of ORF8 are relatively short to have independently foldable structure. It would be really important to conduct some structural analysis of the truncated ORF8 (*e.g.*, by solution NMR). However, such an analysis could be complicated by the presence of two cysteine residues and by the highly hydrophobic nature of this region

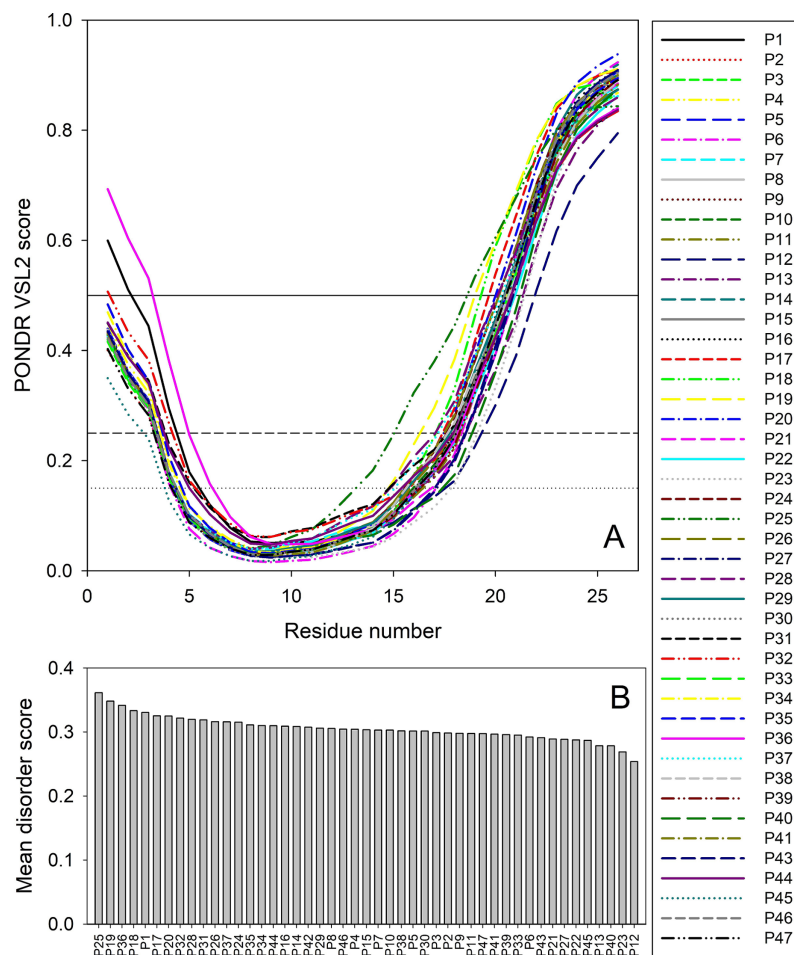


Figure 2 Analysis of intrinsic disorder predisposition of 47 T-ORF8 proteins. (A) Disorder profiles generated using the PONDR-VSL2 disorder predictor. Three thresholds of predicted disorder scores (PDSs) are shown, 0.15, 0.25, and 0.5, which are used for the classification of protein residues as highly disordered ($PDS \geq 0.5$), flexible ($0.25 \leq PDS < 0.5$), moderately flexible ($0.15 \leq PDS < 0.25$) and mostly ordered ($PDS < 0.15$). B. Ranking of 47 T-ORF8 proteins based on their mean disorder scores.

Full-size DOI: [10.7717/peerj.13136/fig-2](https://doi.org/10.7717/peerj.13136/fig-2)

showing the mean hydrophobicity of 0.5926, which dramatically exceeds the mean hydrophobicity of typical globular proteins ($\sim 0.46 \pm 0.05$).

Therefore, to gain some structure-related information for the truncated forms of ORF8, we analyzed the peculiarities of the distribution of per-residue intrinsic disorder predisposition within sequences of the 47 T-ORF8 variants. Since the amino acid sequences of T-ORF8 proteins are shorter than 30 residues, the number of computational tools capable of predicting the intrinsic disorder is limited. In this study, we used the PONDR[®] VSL2 algorithm. The results of this analysis are shown in Fig. 2. Due to their short length and limited sequence variability, T-ORF8 proteins are characterized by rather featureless disorder profiles, where both N- and C-terminal regions are predicted to have higher levels of intrinsic disorder than the central parts.

Most T-ORF8 proteins show rather similar profiles, with the noticeable exceptions of P1 and P36, which show the highest disorder levels in their N-terminal regions. In contrast, P45 presents the least disorder at the N-terminus, P25 the longest and most peculiar disorder distribution in its C-terminal half, P18 and P19 long disorder stretches at their C-termini, and P12 the lowest levels of disorder in the C-terminal region (Fig. 2A). These observations are further supported by Fig. 2B, where the 47 T-ORF8 proteins are ranked based on their mean disorder scores, from the highest to the lowest levels of disorder. However, the vast majority of T-ORF8 proteins (38 of 47) form a rather uniform cluster with the average mean disorder score of 0.304 ± 0.010 , whereas P25, P19, P36, P18, and P1 show higher than average and P13, P40, P23, and P12 lower than average levels of disorder.

Table S3 lists potential functional motifs identified in the 47 T-ORF8 variants by ELM resource and shows that all T-ORF8 proteins have several such motifs. Based on their content of functional motifs, T-ORF8 proteins can be grouped into 21 clusters, with three clusters containing 13, 4, and 2 proteins, respectively, and all the remaining being singletons. The common motif found in all T-ORF8 proteins is the N-degron that initiates protein degradation by binding to the UBR-box of N-recognins. A kinase docking motif that mediates interaction towards the ERK1/2 and P38 subfamilies of MAP kinases and a Ser/Thr (serine/threonine) residue phosphorylated by the Plk1 kinase are present in 20 clusters, whereas 17 clusters also include a site for attachment of a fucose residue to a serine. The lowest number of functional motifs (3) is found in 6 proteins (P12, P16, P17, P19, P21, and P40), many of which are characterized by lower mean disorder scores. On the contrary, proteins with the largest number of functional motifs (6 and 7) typically show higher disorder scores. Table S3 shows that truncation might generate functional T-ORF8 variants (or at least variants possessing functional motifs), and that expected functionality of different T-ORF8 proteins can be quite different. It is clear that the results of this computational analysis should be taken with caution, and the functionality of T-ORF8 requires experimental validation.

Variability and commonality of T-ORF8 variants

In the proceeding section, the quantification of unique T-ORF8 variants using various parameters such as polar/non-polar residue sequence homology, amino acid frequency distributions, amino acid conservation through the Shannon entropy, and physicochemical properties is described.

Polarity based variability of T-ORF8 variants

Each unique T-ORF8 variant possessed a binary polar/non-polar sequence and based on the sequence homology of these sequences, a phylogenetic relationship was obtained (Fig. S4). The number of polar and non-polar residues in the unique T-ORF8 variants was found to be almost balanced (50-50 in percentage). Among 26 residue positions of each T-ORF8 variant with the amino acid length of 26 residues, 14 positions (polar residues at positions 1, 5–7, 13 and non-polar residues at positions 2, 17–18, 20–23) remained

invariant as illustrated by [Table S4](#). The pairs of unique T-ORF8 variants (P5, P6), (P21, P40), (P4, P33), (P12, P13), (P28, P29), and (P9, P10) were closest to each other ([Fig. S4](#)).

Note that, the P15 variant was placed in a single leaf and found to be distant from the other unique ORF8 variants as per polarity-based homology, although P15 was closest to the T-ORF8 variants P13 and P14 based on amino acid homology. Furthermore, it was noticed that only 17 unique T-ORF8 variants possessed unique polar/non-polar sequences ([Table S4](#)). The polar/non-polar sequence of each T-ORF8 variant other than P4, P5, P15, and P28 was unique.

Surprisingly, among the total of 47 T-ORF8 variants, there were 28 T-ORF8 variants, which share identical polar/non-polar sequence with that of P15. According to the phylogenetic relationship derived from the unique polar/non-polar sequence homology, the T-ORF8 P15 was found to be closest to P42. Furthermore, the pairs (P13, P25), (P21, P40), and (P5, P30) were found to be close enough to each other (see [Fig. S5](#)).

Variability of the frequency distribution of amino acids present in T-ORF8 variants

The frequency of each amino acid present in the unique T-ORF8 variants was enumerated, and consequently, a twenty-dimensional frequency vector was obtained ([Table S5](#)).

Tryptophan (Trp) was not present in any of the unique T-ORF8 variants. It was noted that the amino acids arginine (Arg), asparagine (Asn), aspartic acid (Asp), proline (Pro), and tyrosine (Tyr) were absent in the T-ORF8 P15. Arg was found with frequency one in the T-ORF8 P17, P39, P19, P22, and P33. In the P14 and P41 sequences, Asn was present with frequency one. Likewise, Asp was found in P38 and P11. Pro with frequency one, was found in the P8 variant only. Tyr was found in the T-ORF8 P20 and P23 variants. The highest frequency of 4 was seen for phenylalanine (Phe) in P3, P16, and P47, Leu in P10, P12, P24, P34, and P35, and threonine (Thr) in P29, P31, P32, and P42.

For each pair of frequency vectors corresponding to all unique T-ORF8 variants, Euclidean distances were calculated ([Table S6](#)), and the distance matrix in color heat-map is presented in [Fig. 3](#). It was found that the P15 variant is equidistant (1.41) from all other variants except P30 and P40, which are one distance apart from P15. Further, we observed that the distance between any two pairs of T-ORF8 variants is two (light green color) except in a few cases ([Fig. S6](#)). Although the amino acid sequences were different, identical frequency vectors were found for the pair of ORF8 variants (P3, P47), (P2, P9), (P6, P13), (P17, P19), (P24, P34), (P24, P35), (P25, P36), (P34, P35), (P29, P42), and (P27, P43).

Based on the distance matrix, all unique T-ORF8 variants were clustered, and the associated phylogeny is presented in [Fig. S6](#). The P15 variant was very close to P22, P23, P33, P40, and P41 according to the phylogenetic relationship depicted in [Fig. S6](#). Other than the pairs of T-ORF8 having identical frequency vectors, it was found that the pairs of the unique T-ORF8 variants (P23, P33), (P14, P30), (P19, P20), and (P31, P32) were close to each other as derived from the phylogenetic relationship ([Fig. S6](#)).

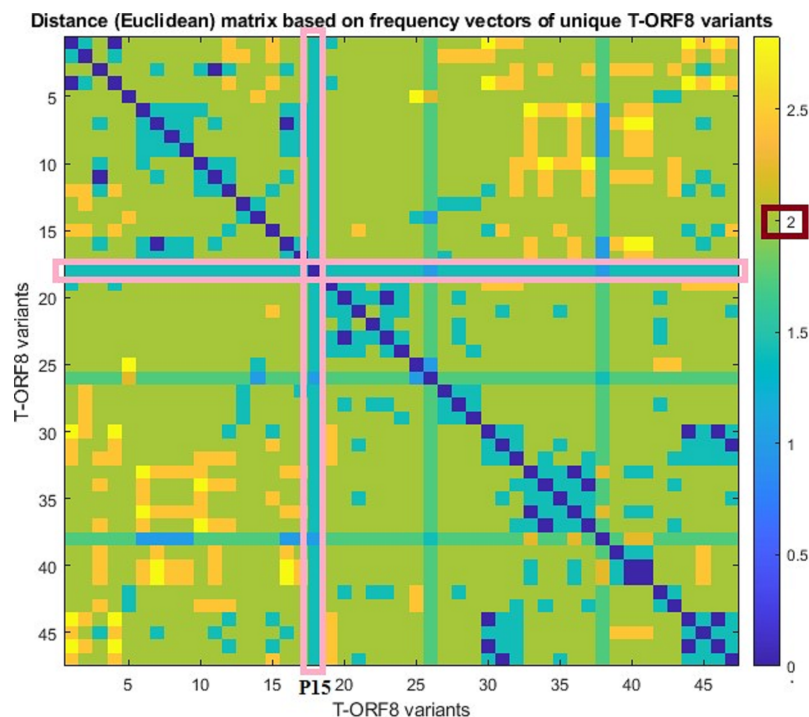


Figure 3 Pairwise distance matrix of amino acid frequency vectors of the unique T-ORF8 variants.

Full-size  DOI: [10.7717/peerj.13136/fig-3](https://doi.org/10.7717/peerj.13136/fig-3)

Variability of T-ORF8 through Shannon entropy

Shannon entropy (SE) for each unique T-ORF8 variant was calculated using the formula stated in “Deep Phylogenetic Analyses” (Table S7). It was found that the highest and lowest SEs of the 47 unique T-ORF8 proteins were 0.958 and 0.973, respectively. That is, the length of the largest interval is 0.015, which is sufficiently small. Based on SEs of the T-ORF8 proteins a set of clusters were derived (Fig. S7A) SEs of each of the T-ORF8 variants are plotted in Fig. S7B. The largest cluster containing 18 T-ORF8 variants (including the T-ORF8 P15) based on the identical SEs were obtained (Fig. S7).

Molecular and physicochemical informatics of unique T-ORF8 variants

For each unique T-ORF8 variant and complete ORF8 protein, several physicochemical and molecular properties were computed using the web-servers. It was found that the extinction coefficient of all T-ORF8 variants was 125, except for four T-ORF8 variants, P16, P17, P18, and P19, whose extinction coefficient was zero (see Table S8). Further, it was noticed that for P20 and P23, extinction coefficients were found to be significantly higher compared to the others. Instability indices of all the T-ORF8 protein variants were ranging from 45.36 to 95.85 (greater than 40), and consequently they are all unstable. It was observed that the P15 variant had a unique frequency of the various types of residues (Tiny: 10, Small: 12, Aliphatic: 9, Aromatic: 4, Non-polar: 16, Polar: 10, Charged: 3, Basic: 2, Acidic: 1) and none of the other T-ORF8 variants was identical.

Furthermore, Euclidean distances between every pair of molecular and physicochemical property vectors corresponding to each T-ORF8 variant were computed and a

phylogenetic relationship was derived (Fig. 4) based on the distance matrix (Table S9). Note that the property vectors of P20 and P30 were highly distant from the other ORF8 variants due to the huge difference in extinction coefficients (for P20, EC: 1490 and for P30, EC: 1615). So, ignoring these two ORF8 variants, the phylogenetic relationship among the remaining 45 T-ORF8 was derived. It was found that none of the T-ORF8 variants had identical property vectors to the P15 variant. It was further found from the phylogenetic relationship (Fig. 4B) that the pair of unique T-ORF8 variants (P17, P18), (P8, P22), (P4, P33), (P28, P29), (P2, P9), (P27, P43), (P7, P39), (P15, P30), (P25, P36), (P26, P45), (P24, P34), (P11, P14), (P21, P40), (P3, P47), and (P31, P32) were found to be the closest pairs based on the closeness of property vectors. Property vector distances from each 45 unique T-ORF8 variant from the P15 variant are presented in Table S10. In the close vicinity of P15, only the P25, P30, and P36 variants appeared based on the nearness of property vectors (Table S10).

Possible T-ORF8 variants in the likelihood of P15 variant

Based on the amino acid sequence homology and other various features, such as the frequency distribution of amino acids, SE, and physicochemical properties of T-ORF8 variants, a possible cluster of nine unique T-ORF8 variants are derived. A schematic presentation is given in Fig. 5. All these nine unique T-ORF8 variants had unique polar/non-polar sequences as shown in Table S4. In addition to the P15 variant, these possible nine emerging variants are likely to appear in the B.1.1.7 lineage of SARS-CoV-2 in the near future. As of May 22nd, 2021, it was observed that 16 of 17 COVID-19 affected patients from India (mostly from Gujrat), were infected by the B.1.1.7 lineage of SARS-CoV-2 with the P15 variant, and only one patient (Accession: QVO43928) infected on February 28, 2021 with the SARS-CoV-2 strain with the P34 T-ORF8 variant, which had an identical polar/non-polar sequence as that of P15.

DISCUSSION AND CONCLUDING REMARKS

The ORF8 protein is 121 amino acid long with two genotypes (orf8L and orf8S). It has an Ig-like fold, is highly immunogenic, and interacts with 47 human proteins, 15 of them being drug targets interacting with MHC-I molecules leading to a significant down-regulation of their surface expression on various cell types (Rashid *et al.*, 2021; Gordon *et al.*, 2020; Gamage *et al.*, 2020). As a result, inhibition of ORF8 could boost special immune surveillance and speed up SARS-CoV-2 eradication *in vivo* (Gamage *et al.*, 2020). ORF8 is not like ORF3a, an ion channel (viroporin), implicated in virion assembly and membrane budding both in SARS-CoV and SARS-CoV-2. In contrast to SARS-CoV, which is not viable when lacking E and ORF3a proteins, and which requires the full-length E and ORF3a proteins for maximal replication and virulence (Barrantes, 2021; Castaño-Rodríguez *et al.*, 2018; Tan *et al.*, 2021), ORF8 in SARS-CoV-2 seems to only have a minor or no impact on the viral life cycle, as the virus can seemingly survive without a functional ORF8 protein, which has been demonstrated by the presence of many mutations and truncations detected in viable SARS-CoV-2 variants (Wang *et al.*, 2020a, 2021). The Q27STOP mutation in the ORF8 protein has been discovered to cause 47

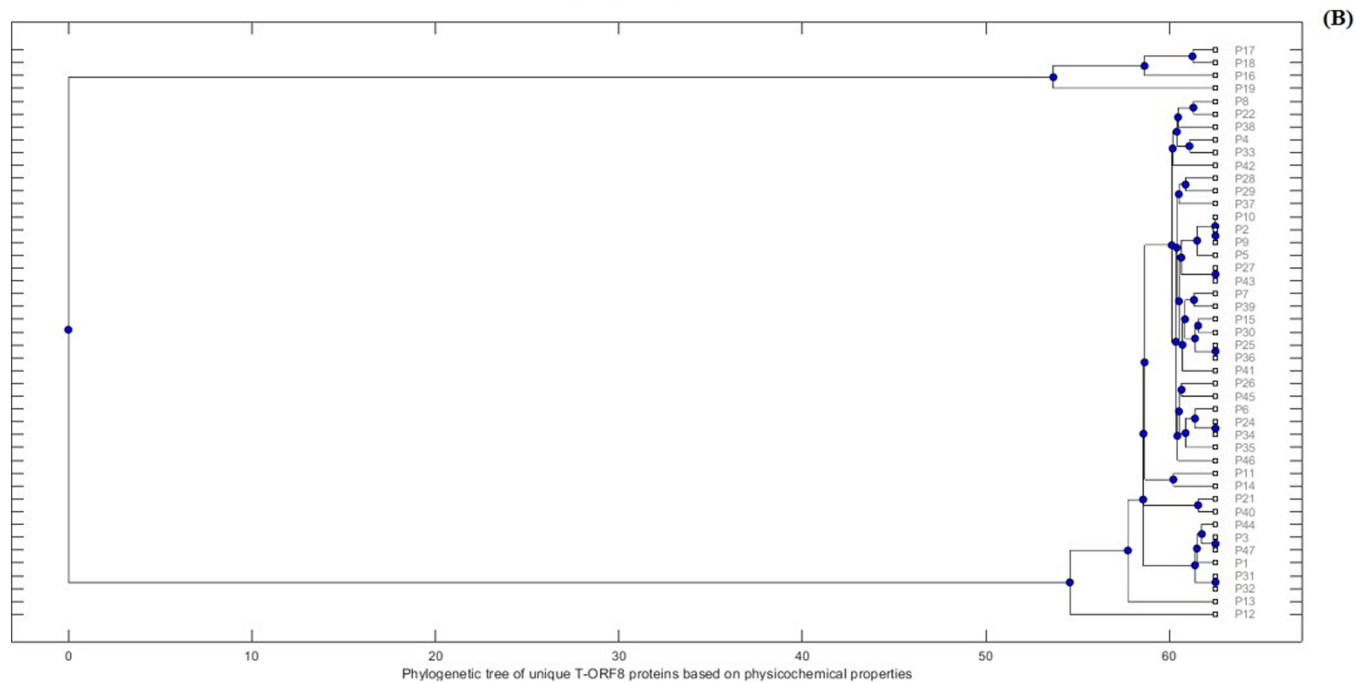
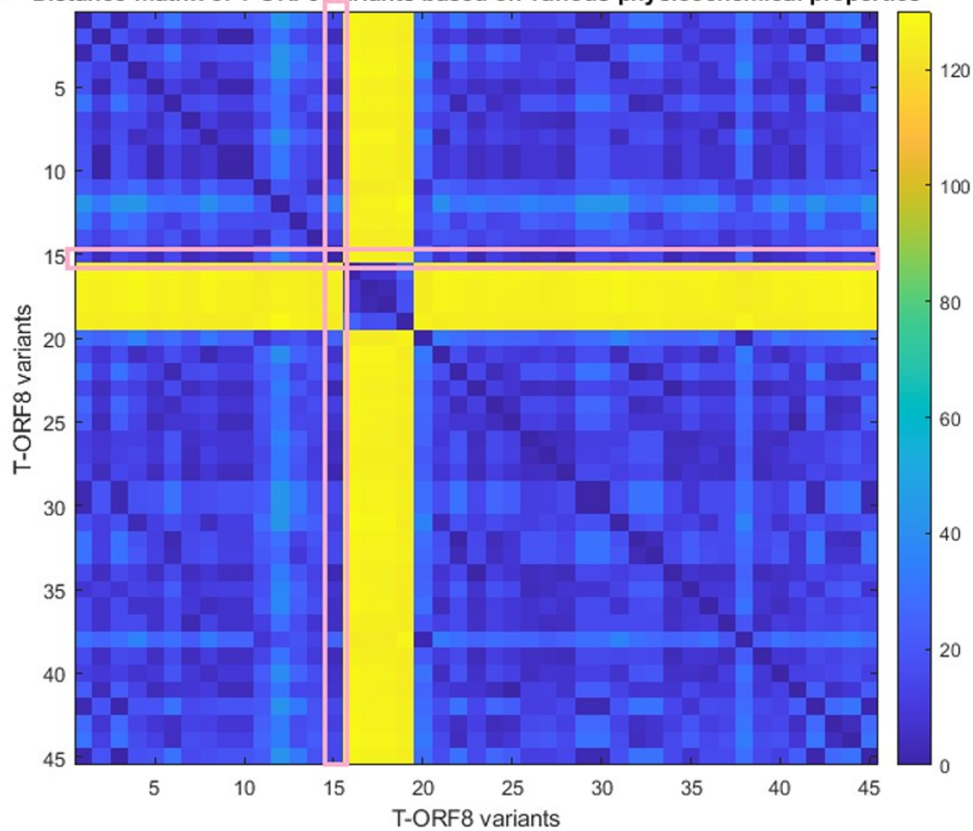
(A) Distance matrix of T-ORF8 variants based on various physicochemical properties

Figure 4 Distance matrix of property vectors and derived phylogenetic tree of 45 T-ORF8 variants. (A) The distance matrix; (B) phylogenetic tree based on physicochemical properties.

Full-size DOI: 10.7717/peerj.13136/fig-4

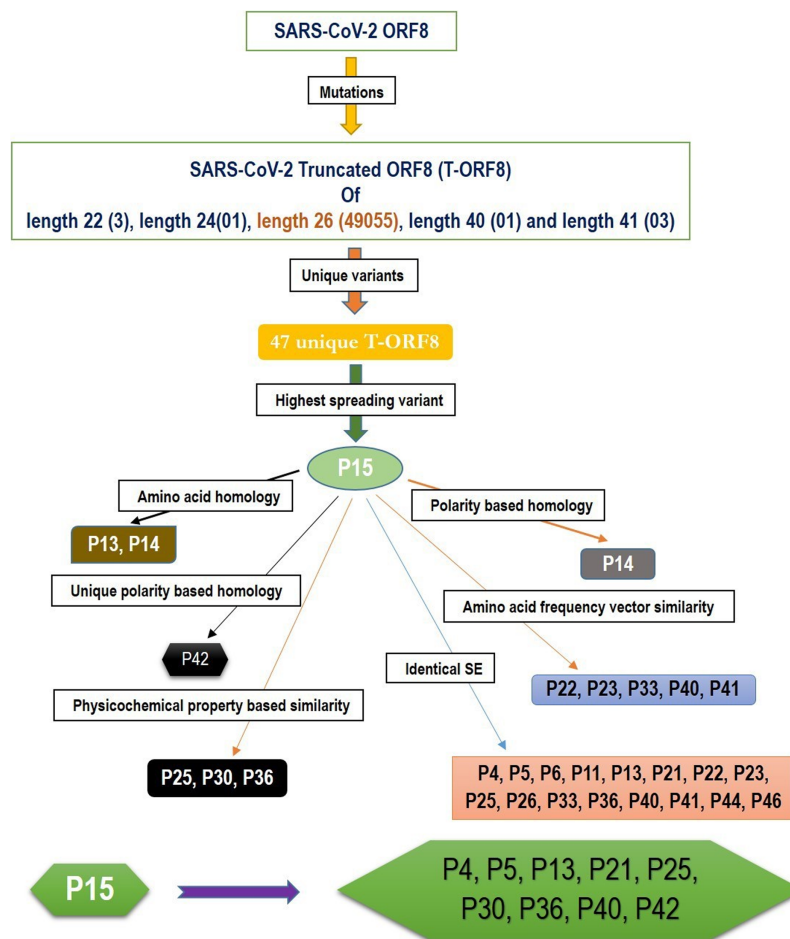


Figure 5 A schematic representation of a possible cluster of unique T-ORF8 variants which were residing in the likelihood of P15 variant. Note: the frequency of each length of T-ORF8 protein was mentioned in parentheses. T-ORF8 variants mentioned in each box were found in the close likelihood of P15 T-ORF8 variants. [Full-size !\[\]\(fcc3264021d438d9732560e78099f674_img.jpg\) DOI: 10.7717/peerj.13136/fig-5](https://doi.org/10.7717/peerj.13136/fig-5)

distinct truncated ORF8 variations. Furthermore, other truncated protein variants of different lengths 22, 24, 40, and 41 amino acids, were detected, although the frequency of occurrences of those variants was significantly lower (Table 2). In Colorado, one T-ORF8 variant of the length of 24 amino acids was detected on April 24, 2021, and this variant is likely to spread further in the future. Other truncated ORF8 variants of amino acids lengths of 22, 40, and 41, no longer appeared in new strains of SARS-CoV-2.

An important consequence of the ORF8 truncation is the alterations in the presentation of this protein by the human leukocyte antigen (HLA) complex on the surface of T-cells, which are among the key players in the immune response to viral infection. In fact, the development of the T-cell-based immunity is based on the presentation of short viral peptides on the cell surface by the HLA complex. Activation of killer CD8 T-cells depends on the recognition of viral peptides presented by HLA class I molecules, whereas activation of CD4 T lymphocytes (helper T-cells) is initiated by binding to a complex between viral peptides and HLA class II molecules (Klein & Sato, 2000). Although the

major functions of the activated CD8 T-cells are the recognition and elimination of the infected cells, the main function of activated CD4 T-cells is regulation of immunity, including stimulation of antibody generation by B cells and enhancement of CD8 T-cell responses (Klein & Sato, 2000; Swain, McKinstry & Strutt, 2012). The overall importance of the T-cell-based responses in COVID-19 severity and long-term immunity has been documented in multiple reports (Wang et al., 2020b; Shkurnikov et al., 2021; Iturrieta-Zuazo et al., 2020; Bange et al., 2021; Grifoni et al., 2020; Peng et al., 2020; Dan et al., 2021). Therefore, it is not surprising that constant appearance of novel mutations in the SARS-CoV-2 genome can target T-cell epitopes (Agerer et al., 2021; Reynolds et al., 2021). Potential CD8 and CD4 epitopes represent peptides derived from the viral proteins, where 8- to 14-mers and 15- to 20-mers are serving as partners for HLA class I and class II receptors, respectively. The appearance of the premature stop codon in the ORF8 gene encoding the NS8 (ORF8) protein (Q27stop) not only generates the truncated form of this protein due to the deletion of almost 80% of the ORF8 protein, but also eliminates the whole immunopeptidome associated with this part of protein. Therefore, it is likely that such distortion in the immunopeptidome of ORF8 will dramatically affect its immunogenicity. Furthermore, mutations found in T-ORF8 could further modulate T-cell immunity.

Quantitative characteristics of the 47 unique truncated ORF8 protein variants were examined. All 47 T-ORF8 variants were found in North America, and only the P15 T-ORF8 variant was spread to four other continents: Africa, Asia, Europe and South America, until May 22, 2021. In this regard, it is pertinent to raise the question of whether there is any correlation between the spreading of all unique T-ORF8 variants and the epidemiological nature of North America. Within North America, it was reported that one of the top mutations, 27964C>T-(S24L) in the ORF8 protein, has an unusually strong gender dependence (Wang et al., 2021). The spread of the P15 variants over the 57 geo-locations across North America was noticed, and in addition, many patients from Asia, Africa, South America and Europe were infected by the particular B.1.1.7 variant of SARS-CoV-2, which contains the P15 variant. Like in many states of the US, also in the US territory of Guam and in North Dakota, most of the patients were noticed to have contracted the P15 variant of the B.1.1.7 lineage. Consequently, the present trend implies that a much higher spread of this lineage with this particular P15 variant is likely to occur. After Europe, the first case of the B.1.1.7 variant with the T-ORF8 P15 was discovered in Maryland in the US, but although later this strain remained limited in Maryland it spread further to other states, such as Florida and Minnesota (Table 3). Furthermore, this analysis reports a set of nine most likely T-ORF8 variants P4, P5, P13, P21, P25, P30, P36, P40, and P42, which were found to be residing in close vicinity of the P15 ORF8 variant. It was noticed that among 47 unique T-ORF8 variants, 28 of them had identical polar/non-polar sequences to that of the P15 variant. Considering the ability of the P15 variant to spread one can assume that the 28 variants with identical polar/non-polar sequences may spread in the near future and cause third, fourth, and fifth waves of COVID-19. As evidence, one patient from India was infected with SARS-CoV-2 with the P34 variant, which has the same polar/non-polar sequences as the P15 variant, as of May

22, 2021 (NCBI accession: QVO43928). The fact that T-ORF8 is still operating as ORF8, is an open issue that needs to be addressed. Reports try to link these T-ORF8 present in many lineages to COVID-19 severity and/or outcomes, effects that contribute to disease progression if associated with mutations in spike protein ([Pereira, 2020](#); [Guthmiller et al., 2021](#); [Nagy, Pongor & Györfy, 2021](#)). It has also been reported that patients infected with SARS-CoV-2, lacking the majority of ORF8 protein, have a lower risk of aggravation, a conclusion that accrued variants in the spike, ORF8, and ORF3a proteins were associated with improved clinical outcomes ([Esper et al., 2021](#); [Young et al., 2020](#)). More recently, SARS-CoV-2 strains were isolated in Washington state (USA), with a stop codon mutation generating a novel truncated and much shorter ORF8 protein, as well as in Hong Kong, which completely lacked ORF8 (gene, protein and antibody), ORF7a, and ORF7b ([Esper et al., 2021](#); [DeRonde et al., 2021](#); [Tse et al., 2021](#)). However, the *in vitro* analysis on Nasal Epithelial cells (NECs) infected with one of these isolates (ORF8-Δ382) showed no significant functional differences between the wild type ORF8 and the ORF-Δ382 mutant ([Gamage et al., 2020](#)). In contrast, Vero-6 cell infected with the same strain (ORF8-Δ382) showed significantly higher replicative fitness *in vitro* than the wild type, while no difference was observed in the patient viral load, indicating that the deletion variant retained its replicative fitness ([Su et al., 2020](#)). In any case, the combinatorial clinical effects of T-ORF8 need to be investigated and analyzed in depth. It is necessary to investigate in detail the functions of T-ORF8 and the effects of this protein on inflammation and antigen-presenting ability. Finally, caution should be paid to using ORF8 as a diagnostic marker, as many immunoassay tests depend on antibody potency ([Pereira, 2021](#)). A systematic analysis of this protein and its specific antibodies is needed to determine the effects of these mutations/truncations on the diagnostic potential of the anti-ORF8 antibodies.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

Vasco Azevedo, Debmalya Barh and Vladimir N Uversky are Academic Editors for PeerJ. Kenneth Lundstrom is employed by PanTherapeutics.

Author Contributions

- Sk. Sarif Hassan conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Vaishnavi Kodakandla performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Elrashdy M. Redwan performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

- Kenneth Lundstrom conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Pabitra Pal Choudhury performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Tarek Mohamed Abd El-Aziz performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Kazuo Takayama performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Ramesh Kandimalla performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Amos Lal performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Ángel Serrano-Aroca performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Gajendra Kumar Azad performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Alaa AA. Aljabali performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Giorgio Palù performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Gaurav Chauhan performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Parise Adadi performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Murtaza Tambuwala performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Adam M. Brufsky performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Wagner Baetas-da-Cruz performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Debmalya Barh performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Vasco Azevedo performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Nikolas G. Bazan performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Bruno Silva Andrade conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Raner José Santana Silva performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

- Vladimir N. Uversky conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data are available in the article and the [Supplemental File](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.13136#supplemental-information>.

REFERENCES

- Agerer B, Koblischke M, Gudipati V, Montaña-Gutierrez LF, Smyth M, Popa A, Genger JW, Endler L, Florian DM, Mühlgrabner V, Graninger M, Aberle SW, Husa AM, Shaw LE, Lercher A, Gattinger P, Torralba-Gombau R, Trapin D, Penz T, Barreca D, Fae I, Wenda S, Traugott M, Walder G, Pickl WF, Thiel V, Allerberger F, Stockinger H, Puchhammer-Stöckl E, Weninger W, Fischer G, Hoepler W, Pawelka E, Zoufaly A, Valenta R, Bock C, Paster W, Geyeregger R, Farlik M, Halbritter F, Huppa JB, Aberle JH, Bergthaler A. 2021. SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8 + T cell responses. *Science Immunology* 6(57):eabg6461 DOI 10.1126/sciimmunol.abg6461.
- Alkhansa A, Lakkis G, El Zein L. 2021. Mutational analysis of SARS-CoV-2 ORF8 during six months of COVID-19 pandemic. *Gene Reports* 23:101024 DOI 10.1016/j.genrep.2021.101024.
- Bange EM, Han NA, Wileyto P, Kim JY, Gouma S, Robinson J, Greenplate AR, Hwee MA, Porterfield F, Owoyemi O, Naik K, Zheng C, Galantino M, Weisman AR, Ittner CAG, Kugler EM, Baxter AE, Oniyide O, Agyekum RS, Dunn TG, Jones TK, Giannini HM, Weirick ME, McAllister CM, Babady NE, Kumar A, Widman AJ, DeWolf S, Boutemine SR, Roberts C, Budzik KR, Tollett S, Wright C, Perloff T, Sun L, Mathew D, Giles JR, Oldridge DA, Wu JE, Alanio C, Adamski S, Garfall AL, Vella LA, Kerr SJ, Cohen JV, Oyer RA, Massa R, Maillard IP, Maxwell KN, Reilly JP, Maslak PG, Vonderheide RH, Wolchok JD, Hensley SE, Wherry EJ, Meyer NJ, DeMichele AM, Vardhana SA, Mamtani R, Huang AC. 2021. CD8+ T cells contribute to survival in patients with COVID-19 and hematologic cancer. *Nature Medicine* 27(7):1280–1289 DOI 10.1038/s41591-021-01386-7.
- Barrantes FJ. 2021. Structural biology of coronavirus ion channels. *Acta Crystallographica Section D: Structural Biology* 77(Pt 4):391–402 DOI 10.1107/S2059798321001431.
- Benvenuto D, Giovanetti M, Salemi M, Prosperi M, De Flora C, Junior Alcantara LC, Angeletti S, Ciccozzi M. 2020. The global spread of 2019-nCoV: a molecular evolutionary analysis. *Pathogens and Global Health* 114(2):64–67 DOI 10.1080/20477724.2020.1725339.
- Borges V, Sousa C, Menezes L, Gonçalves AM, Picão M, Almeida JP, Vieita M, Santos R, Silva AR, Costa M, Carneiro L. 2021. Tracking SARS-CoV-2 VOC 202012/01 (lineage B. 1.1. 7) dissemination in Portugal: insights from nationwide RT-PCR Spike gene drop out data. *Eurosurveillance* 26:2100131 DOI 10.2807/1560-7917.ES.2021.26.10.2100130.
- Broome BM, Hecht MH. 2000. Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. *Journal of Molecular Biology* 296(4):961–968 DOI 10.1006/jmbi.2000.3514.

- Castaño-Rodríguez C, Honrubia JM, Gutiérrez-Álvarez J, DeDiego ML, Nieto-Torres JL, Jimenez-Guardeño JM, Regla-Nava JA, Fernandez-Delgado R, Verdia-Báguena C, Queralt-Martín M, Kochan G, Perlman S, Aguilera VM, Sola I, Enjuanes L. 2018. Role of severe acute respiratory syndrome coronavirus viroporins E, 3a, and 8a in replication and pathogenesis. *mBio* 9(3):e02325 DOI 10.1128/mBio.02325-17.
- Chasapis CT, Georgiopoulou AK, Perlepes SP, Björklund G, Peana M. 2021. A SARS-CoV-2-human metalloproteome interaction map. *Journal of Inorganic Biochemistry* 219:111423 DOI 10.1016/j.jinorgbio.2021.111423.
- Colson P, Delerce J, Burel E, Dahan J, Jouffret A, Fenollar F, Yahi N, Fantini J, La Scola B, Raoult D. 2021. Emergence in Southern France of a new SARS-CoV-2 variant of probably Cameroonian origin harbouring both substitutions N501Y and E484K in the spike protein. *medRxiv* DOI 10.1101/2021.12.24.21268174.
- Dan JM, Mateus J, Kato Y, Hastie KM, Yu ED, Faliti CE, Grifoni A, Ramirez SI, Haupt S, Frazier A, Nakao C, Rayaprolu V, Rawlings SA, Peters B, Krammer F, Simon V, Sapphire EO, Smith DM, Weiskopf D, Sette A, Crotty S. 2021. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* 371(6529):eabf4063 DOI 10.1126/science.abf4063.
- Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, Wenseleers T, Gimma A, Waites W, Wong KLM, van Zandvoort K, Silverman JD, CMMID COVID-19 Working Group; COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz K, Keogh R, Eggo RM, Funk S, Jit M, Atkins KE, Edmunds WJ. 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372(6538):eabg3055 DOI 10.1126/science.abg3055.
- DeRonde S, Deuling H, Parker J, Chen J. 2021. Identification of a novel SARS-CoV-2 strain with truncated protein in ORF8 gene by next generation sequencing. *Research Square* rs.3.rs:413141 DOI 10.21203/rs.3.rs-413141/v1.
- Díaz J. 2020. SARS-CoV-2 molecular network structure. *Frontiers in Physiology* 11:870 DOI 10.3389/fphys.2020.00870.
- Esper FP, Cheng YW, Adhikari TM, Tu ZJ, Li D, Li EA, Farkas DH, Procop GW, Ko JS, Chan TA, Jehi L, Rubin BP, Li J. 2021. Genomic epidemiology of SARS-CoV-2 infection during the initial pandemic wave and association with disease severity. *JAMA Network Open* 4(4):e217746 DOI 10.1001/jamanetworkopen.2021.7746.
- Flower TG, Buffalo CZ, Hooy RM, Allaire M, Ren X, Hurley JH. 2021. Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. *Proceedings of the National Academy of Sciences of the United States of America* 118(2):e2021785118 DOI 10.1073/pnas.2021785118.
- Fontanet A, Autran B, Lina B, Kieny MP, Karim SSA, Sridhar D. 2021. SARS-CoV-2 variants and ending the COVID-19 pandemic. *Lancet* 397(10278):952–954 DOI 10.1016/S0140-6736(21)00370-6.
- Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, Scott R, Sconza R, Price J, Margaritis M, Bergstrom M, Spyer MJ, Miralhes PB, Grant P, Kirk S, Valerio C, Mangera Z, Prabhakar T, Moreno-Cuesta J, Arulkumaran N, Singer M, Shin GY, Sanchez E, Paraskevopoulou SM, Pillay D, McKendry RA, Mirfenderesky M, Houlihan CF, Nastouli E. 2021. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *Lancet Infectious Diseases* 21(9):1246–1256 DOI 10.1016/S1473-3099(21)00170-5.
- Galloway SE, Paul P, MacCannell DR, Johansson MA, Brooks JT, MacNeil A, Slayton RB, Tong S, Silk BJ, Armstrong GL, Biggerstaff M, Dugan VG. 2021. Emergence of SARS-CoV-2

B.1.1.7 Lineage—United States, December 29, 2020–January 12, 2021. *MMWR—Morbidity and Mortality Weekly Report* 70(3):95–99 DOI 10.15585/mmwr.mm7003e2.

- Gamage AM, Tan KS, Chan WOY, Liu J, Tan CW, Ong YK, Thong M, Andiappan AK, Anderson DE, Wang Y, Wang LF. 2020. Infection of human Nasal Epithelial Cells with SARS-CoV-2 and a 382-nt deletion isolate lacking ORF8 reveals similar viral kinetics and host transcriptional profiles. *PLoS Pathogens* 16(12):e1009130 DOI 10.1371/journal.ppat.1009130.
- Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook*. Totowa: Humana Press, 571–607.
- Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O’Meara MJ, Rezelj VV, Guo JZ, Swaney DL, Tummino TA, Hüttenhain R, Kaake RM, Richards AL, Tutuncuoglu B, Foussard H, Batra J, Haas K, Modak M, Kim M, Haas P, Polacco BJ, Braberg H, Fabius JM, Eckhardt M, Soucheray M, Bennett MJ, Cakir M, McGregor MJ, Li Q, Meyer B, Roesch F, Vallet T, Mac Kain A, Miorin L, Moreno E, Naing ZZC, Zhou Y, Peng S, Shi Y, Zhang Z, Shen W, Kirby IT, Melnyk JE, Chorba JS, Lou K, Dai SA, Barrio-Hernandez I, Memon D, Hernandez-Armenta C, Lyu J, Mathy CJP, Perica T, Pilla KB, Ganesan SJ, Saltzberg DJ, Rakesh R, Liu X, Rosenthal SB, Calviello L, Venkataramanan S, Liboy-Lugo J, Lin Y, Huang XP, Liu Y, Wankowicz SA, Bohn M, Safari M, Ugur FS, Koh C, Savar NS, Tran QD, Shengjuler D, Fletcher SJ, O’Neal MC, Cai Y, Chang JCJ, Broadhurst DJ, Klippsten S, Sharp PP, Wenzell NA, Kuzuoglu-Ozturk D, Wang HY, Trenker R, Young JM, Cavero DA, Hiatt J, Roth TL, Rathore U, Subramanian A, Noack J, Hubert M, Stroud RM, Frankel AD, Rosenberg OS, Verba KA, Agard DA, Ott M, Emerman M, Jura N, von Zastrow M, Verdin E, Ashworth A, Schwartz O, d’Enfert C, Mukherjee S, Jacobson M, Malik HS, Fujimori DG, Ideker T, Craik CS, Floor SN, Fraser JS, Gross JD, Sali A, Roth BL, Ruggero D, Taunton J, Kortemme T, Beltrao P, Vignuzzi M, García-Sastre A, Shokat KM, Shoichet BK, Krogan NJ. 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583(7816):459–468 DOI 10.1038/s41586-020-2286-9.
- Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, Rawlings SA, Sutherland A, Premkumar L, Jadi RS, Marrama D, de Silva AM, Frazier A, Carlin AF, Greenbaum JA, Peters B, Krammer F, Smith DM, Crotty S, Sette A. 2020. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* 181(7):1489–1501.e15 DOI 10.1016/j.cell.2020.05.015.
- Guthmiller JJ, Stovicek O, Wang J, Changrob S, Li L, Halfmann P, Zheng NY, Utset H, Stamper CT, Dugan HL, Miller WD, Huang M, Dai YN, Nelson CA, Hall PD, Jansen M, Shanmugarajah K, Donington JS, Krammer F, Fremont DH, Joachimiak A, Kawaoka Y, Tesic V, Madariaga ML, Wilson PC. 2021. SARS-CoV-2 infection severity is linked to superior humoral immunity against the spike. *mBio* 12(1):e02940 DOI 10.1128/mBio.02940-20.
- Hachim A, Kavian N, Cohen CA, Chin AWH, Chu DKW, Mok CKP, Tsang OTY, Yeung YC, Perera RAPM, Poon LLM, Peiris JSM, Valkenburg SA. 2020. ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. *Nature Immunology* 21(10):1293–1301 DOI 10.1038/s41590-020-0773-7.
- Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22(2):160–174 DOI 10.1007/BF02101694.
- Hassan SS, Aljabali AAA, Panda PK, Ghosh S, Attrish D, Choudhury PP, Seyran M, Pizzol D, Adadi P, Abd El-Aziz TM, Soares A, Kandimalla R, Lundstrom K, Lal A, Azad GK, Uversky VN, Sherchan SP, Baetas-da-Cruz W, Uhal BD, Rezaei N, Chauhan G, Barh D, Redwan EM, Dayhoff GW II, Bazan NG, Serrano-Aroca Á, El-Demerdash A, Mishra YK, Palu G, Takayama K, Brufsky AM, Tambuwala MM. 2021a. A unique view of SARS-CoV-2

- through the lens of ORF8 protein. *Computers in Biology and Medicine* **133(8)**:104380 DOI [10.1016/j.combiomed.2021.104380](https://doi.org/10.1016/j.combiomed.2021.104380).
- Hassan SS, Attrish D, Ghosh S, Choudhury PP, Roy B. 2021b. Pathogenic perspective of missense mutations of ORF3a protein of SARS-CoV-2. *Virus Research* **300(1)**:198441 DOI [10.1016/j.virusres.2021.198441](https://doi.org/10.1016/j.virusres.2021.198441).
- Hassan SS, Attrish D, Ghosh S, Choudhury PP, Uversky VN, Aljabali AAA, Lundstrom K, Uhal BD, Rezaei N, Seyran M, Pizzol D, Adadi P, Soares A, Abd El-Aziz TM, Kandimalla R, Tambuwala MM, Azad GK, Sherchan SP, Baetas-da-Cruz W, Lal A, Palù G, Takayama K, Serrano-Aroca Á, Barh D, Brufsky AM. 2021c. Notable sequence homology of the ORF10 protein introspects the architecture of SARS-CoV-2. *International Journal of Biological Macromolecules* **181(5)**:801–809 DOI [10.1016/j.ijbiomac.2021.03.199](https://doi.org/10.1016/j.ijbiomac.2021.03.199).
- Hassan SS, Choudhury PP, Basu P, Jana SS. 2020a. Molecular conservation and differential mutation on ORF3a gene in Indian SARS-CoV2 genomes. *Genomics* **112(5)**:3226–3237 DOI [10.1016/j.ygeno.2020.06.016](https://doi.org/10.1016/j.ygeno.2020.06.016).
- Hassan SS, Choudhury PP, Roy B. 2021. Rare mutations in the accessory proteins ORF6, ORF7b, and ORF10 of the SARS-CoV-2 genomes. *Meta Gene* **28(18)**:100873 DOI [10.1016/j.mgene.2021.100873](https://doi.org/10.1016/j.mgene.2021.100873).
- Hassan SS, Choudhury PP, Roy B, Jana SS. 2020b. Missense mutations in SARS-CoV2 genomes from Indian patients. *Genomics* **112(6)**:4622–4627 DOI [10.1016/j.ygeno.2020.08.021](https://doi.org/10.1016/j.ygeno.2020.08.021).
- Hassan SS, Ghosh S, Attrish D, Choudhury PP, Aljabali AAA, Uhal BD, Lundstrom K, Rezaei N, Uversky VN, Seyran M, Pizzol D, Adadi P, Soares A, El-Aziz TMA, Kandimalla R, Tambuwala MM, Azad GK, Sherchan SP, Baetas-da-Cruz W, Takayama K, Serrano-Aroca Á, Chauhan G, Palu G, Brufsky AM. 2020c. Possible transmission flow of SARS-CoV-2 based on ACE2 features. *Molecules* **25(24)**:5906 DOI [10.3390/molecules25245906](https://doi.org/10.3390/molecules25245906).
- Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. 2017. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **33(19)**:3098–3100 DOI [10.1093/bioinformatics/btx345](https://doi.org/10.1093/bioinformatics/btx345).
- Hu B, Guo H, Zhou P, Shi ZL. 2021. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* **19(3)**:141–154 DOI [10.1038/s41579-020-00459-7](https://doi.org/10.1038/s41579-020-00459-7).
- Iturrieta-Zuazo I, Rita CG, García-Soidán A, de Malet Pintos-Fonseca A, Alonso-Alarcón N, Pariente-Rodríguez R, Tejeda-Velarde A, Serrano-Villar S, Castañer-Alabau JL, Nieto-Gañán I. 2020. Possible role of HLA class-I genotype in SARS-CoV-2 infection and progression: a pilot study in a cohort of COVID-19 Spanish patients. *Clinical Immunology* **219(13)**:108572 DOI [10.1016/j.clim.2020.108572](https://doi.org/10.1016/j.clim.2020.108572).
- Klein J, Sato A. 2000. The HLA system. First of two parts. *New England Journal of Medicine* **343(10)**:702–709 DOI [10.1056/NEJM200009073431006](https://doi.org/10.1056/NEJM200009073431006).
- Koyama T, Platt D, Parida L. 2020. Variant analysis of SARS-CoV-2 genomes. *Bulletin of the World Health Organization* **98(7)**:495–504 DOI [10.2471/BLT.20.253591](https://doi.org/10.2471/BLT.20.253591).
- Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J, Diakogianni A, Valverde JA, Bukirova D, Čalyševa J, Palopoli N, Davey NE, Chemes LB, Gibson TJ. 2020. ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Research* **48(D1)**:D296–D306 DOI [10.1093/nar/gkz1030](https://doi.org/10.1093/nar/gkz1030).
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35(6)**:1547–1549 DOI [10.1093/molbev/msy096](https://doi.org/10.1093/molbev/msy096).

- Li JY, Liao CH, Wang Q, Tan YJ, Luo R, Qiu Y, Ge XY. 2020. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. *Virus Research* 286(1):198074 DOI 10.1016/j.virusres.2020.198074.
- Li R, Ma X, Deng J, Chen Q, Liu W, Peng Z, Qiao Y, Lin Y, He X, Zhang H. 2021. Differential efficiencies to neutralize the novel mutants B.1.1.7 and 501Y.V2 by collected sera from convalescent COVID-19 patients and RBD nanoparticle-vaccinated rhesus macaques. *Cellular & Molecular Immunology* 18(4):1058–1060 DOI 10.1038/s41423-021-00641-8.
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47(W1):W636–W641 DOI 10.1093/nar/gkz268.
- Matheson NJ, Lehner PJ. 2020. How does SARS-CoV-2 cause COVID-19? *Science* 369(6503):510–511 DOI 10.1126/science.abc6156.
- Mercatelli D, Giorgi FM. 2020. Geographic and genomic distribution of SARS-CoV-2 mutations. *Frontiers in Microbiology* 11:1800 DOI 10.3389/fmicb.2020.01800.
- Mohammad S, Bouchama A, Mohammad Alharbi B, Rashid M, Saleem Khatlani T, Gaber NS, Malik SS. 2020. SARS-CoV-2 ORF8 and SARS-CoV ORF8ab: genomic divergence and functional convergence. *Pathogens* 9(9):677 DOI 10.3390/pathogens9090677.
- Nagy Á, Pongor S, Györffy B. 2021. Different mutations in SARS-CoV-2 associate with severe and mild outcome. *International Journal of Antimicrobial Agents* 57(2):106272 DOI 10.1016/j.ijantimicag.2020.106272.
- Necci M, Piovesan D, CAID Predictors; DisProt Curators, Tosatto SCE. 2021. Critical assessment of protein intrinsic disorder prediction. *Nature Methods* 18(5):472–481 DOI 10.1038/s41592-021-01117-3.
- Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. 2005. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins-structure Function and Bioinformatics* 61(Suppl 7):176–182 DOI 10.1002/(ISSN)1097-0134.
- Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, Dejnirattisai W, Rostron T, Supasa P, Liu C, López-Camacho C, Slon-Campos J, Zhao Y, Stuart DI, Paesen GC, Grimes JM, Antson AA, Bayfield OW, Hawkins DEDP, Ker DS, Wang B, Turtle L, Subramaniam K, Thomson P, Zhang P, Dold C, Ratcliff J, Simmonds P, de Silva T, Sopp P, Wellington D, Rajapaksa U, Chen YL, Salio M, Napolitani G, Paes W, Borrow P, Kessler BM, Fry JW, Schwabe NF, Semple MG, Baillie JK, Moore SC, Openshaw PJM, Ansari MA, Dunachie S, Barnes E, Frater J, Kerr G, Goulder P, Lockett T, Levin R, Zhang Y, Jing R, Ho LP, Cornall RJ, Conlon CP, Klenerman P, Screaton GR, Mongkolsapaya J, McMichael A, Knight JC, Ogg G, Dong T. 2020. Oxford Immunology network COVID-19 response T cell consortium; ISARIC4C investigators, broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nature Immunology* 21(11):1336–1345 DOI 10.1038/s41590-020-0782-6.
- Perchetti GA, Zhu H, Mills MG, Shrestha L, Wagner C, Bakhsh SM, Lin MJ, Xie H, Huang ML, Mathias P, Bedford T, Jerome KR, Greninger AL, Roychoudhury P. 2021. Specific allelic discrimination of N501Y and other SARS-CoV-2 mutations by ddPCR detects B.1.1.7 lineage in Washington State. *Journal of Medical Virology* 93(10):5931–5941 DOI 10.1002/jmv.27155.
- Pereira F. 2020. Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. *Infection Genetics and Evolution* 85:104525 DOI 10.1016/j.meegid.2020.104525.
- Pereira F. 2021. SARS-CoV-2 variants lacking a functional ORF8 may reduce accuracy of serological testing. *Journal of Immunological Methods* 488:112906 DOI 10.1016/j.jim.2020.112906.

- Ramírez JD, Muñoz M, Patiño LH, Ballesteros N, Paniz-Mondolfi A. 2021. Will the emergent SARS-CoV2 B.1.1.7 lineage affect molecular diagnosis of COVID-19? *Journal of Medical Virology* 93(5):2566–2568 DOI 10.1002/jmv.26823.
- Rashid F, Dzakah EE, Wang H, Tang S. 2021. The ORF8 protein of SARS-CoV-2 induced endoplasmic reticulum stress and mediated immune evasion by antagonizing production of interferon beta. *Virus Research* 296:198350 DOI 10.1016/j.virusres.2021.198350.
- Ren Y, Shu T, Wu D, Mu J, Wang C, Huang M, Han Y, Zhang XY, Zhou W, Qiu Y, Zhou X. 2020. The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cellular & Molecular Immunology* 17(8):881–883 DOI 10.1038/s41423-020-0485-9.
- Reynolds CJ, Pade C, Gibbons JM, Butler DK, Otter AD, Menacho K, Fontana M, Smit A, Sackville-West JE, Cutino-Moguel T, Maini MK, Chain B, Noursadeghi M, Brooks T, Semper A, Manisty C, Treibel TA, Moon JC, Valdes AM, McKnight Áine, Altmann DM, Boyton R, Abbass H, Abiodun A, Alfarihi M, Alldis Z, Altmann DM, Amin OE, Andiapen M, Artico J, Augusto Jão B, Baca GL, Bailey SNL, Bhuva AN, Boulter A, Bowles R, Boyton RJ, Bracken OV, O'Brien B, Brooks T, Bullock N, Butler DK, Captur G, Champion N, Chan C, Chandran A, Collier D, Couto de Sousa J, Couto-Parada X, Cutino-Moguel T, Davies RH, Douglas B, Di Genova C, Dieobi-Anene K, Diniz MO, Ellis A, Feehan K, Finlay M, Fontana M, Forooghi N, Gaier C, Gibbons JM, Gilroy D, Hamblin M, Harker G, Hewson J, Heywood W, Hickling LM, Hingorani AD, Howes L, Hughes A, Hughes G, Hughes R, Itua I, Jardim V, Lee W-YJ, Jensen M, Jones J, Jones M, Joy G, Kapil V, Kurdi H, Lambourne J, Lin K-M, Louth S, Maini MK, Mandadapu V, Manisty C, McKnight Áine, Menacho K, Mfuko C, Mills K, Mitchelmore O, Moon C, Moon JC, Munoz-Sandoval D, Murray SM, Noursadeghi M, Otter A, Pade C, Palma S, Parker R, Patel K, Pawarova B, Petersen SE, Piniera B, Pieper FP, Pope D, Prossora M, Rannigan L, Rapala A, Reynolds CJ, Richards A, Robathan M, Rosenheim J, Sambile G, Schmidt NM, Semper A, Seraphim A, Simion M, Smit A, Sugimoto M, Swadling L, Taylor S, Temperton N, Thomas S, Thornton GD, Treibel TA, Tucker A, Veerapen J, Vijayakumar M, Welch S, Wodehouse T, Wynne L, Zahedi D, Chain B. 2021. Prior SARS-CoV-2 infection rescues B and T cell responses to variants after first vaccine dose. *Science* 30(6549):1418–1423 DOI 10.1126/science.abh1282.
- Sengupta A, Hassan SS, Choudhury PP. 2021. Clade GR and clade GH isolates of SARS-CoV-2 in Asia show highest amount of SNPs. *Infection Genetics and Evolution* 89(5):104724 DOI 10.1016/j.meegid.2021.104724.
- Shen X, Tang H, McDanal C, Wagh K, Fischer W, Theiler J, Yoon H, Li D, Haynes BF, Sanders KO, Gnanakaran S, Hengartner N, Pajon R, Smith G, Glenn GM, Korber B, Montefiori DC. 2021. SARS-CoV-2 variant B.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines. *Cell Host & Microbe* 29(4):529–539.e3 DOI 10.1016/j.chom.2021.03.002.
- Shkurnikov M, Nersisyan S, Jankevic T, Galatenko A, Gordeev I, Vechorko V, Tonevitsky A. 2021. Association of HLA class I genotypes with severity of coronavirus disease-19. *Frontiers in Immunology* 12:641900 DOI 10.3389/fimmu.2021.641900.
- Stecher G, Tamura K, Kumar S. 2020. Molecular evolutionary genetics analysis (MEGA) for macOS. *Molecular Biology and Evolution* 37(4):1237–1239 DOI 10.1093/molbev/msz312.
- Strait BJ, Dewey TG. 1996. The Shannon information entropy of protein sequences. *Biophysical Journal* 71(1):148–155 DOI 10.1016/S0006-3495(96)79210-X.
- Stukalov A, Girault V, Grass V, Karayel O, Bergant V, Urban C, Haas DA, Huang Y, Oubraham L, Wang A, Hamad MS, Piras A, Hansen FM, Tanzer MC, Paron I, Zinzula L, Engleitner T, Reinecke M, Lavacca TM, Ehmann R, Wölfel R, Jores J, Kuster B, Protzer U, Rad R, Ziebuhr J, Thiel V, Scaturro P, Mann M, Pichlmair A. 2021. Multilevel proteomics

- reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594**(7862):246–252
DOI [10.1038/s41586-021-03493-4](https://doi.org/10.1038/s41586-021-03493-4).
- Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, Zhuang Y, Kalimuddin S, Low JGH, Tan CW, Chia WN, Mak TM, Octavia S, Chavatte JM, Lee RTC, Pada S, Tan SY, Sun L, Yan GZ, Maurer-Stroh S, Mendenhall IH, Leo YS, Lye DC, Wang LF, Smith GJD. 2020.** Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. *mBio* **11**(4):e01610 DOI [10.1128/mBio.01610-20](https://doi.org/10.1128/mBio.01610-20).
- Swain SL, McKinsty KK, Strutt TM. 2012.** Expanding roles for CD4+ T cells in immunity to viruses. *Nature Reviews Immunology* **12**(2):136–148 DOI [10.1038/nri3152](https://doi.org/10.1038/nri3152).
- Tan Y, Schneider T, Shukla PK, Chandrasekharan MB, Aravind L, Zhang D. 2021.** Unification and extensive diversification of M/Orf3-related ion channel proteins in coronaviruses and other nidoviruses. *Virus Evolution* **7**(1):veab014 DOI [10.1093/ve/veab014](https://doi.org/10.1093/ve/veab014).
- Tse H, Lung DC, Wong SC, Ip KF, Wu TC, To KK, Kok KH, Yuen KY, Choi GK. 2021.** Emergence of a severe acute respiratory syndrome coronavirus 2 virus variant with novel genomic architecture in Hong Kong. *Clinical Infectious Diseases* **73**(9):1696–1699 DOI [10.1093/cid/ciab198](https://doi.org/10.1093/cid/ciab198).
- Velazquez-Salinas L, Zarate S, Eberl S, Gladue DP, Novella I, Borca MV. 2020.** Positive selection of ORF1ab, ORF3a, and ORF8 genes drives the early evolutionary trends of SARS-CoV-2 during the 2020 COVID-19 pandemic. *Frontiers in Microbiology* **11**:550674 DOI [10.3389/fmicb.2020.550674](https://doi.org/10.3389/fmicb.2020.550674).
- Walensky RP, Walke HT, Fauci AS. 2021.** SARS-CoV-2 variants of concern in the United States—challenges and opportunities. *JAMA* **325**(11):1037–1038 DOI [10.1001/jama.2021.2294](https://doi.org/10.1001/jama.2021.2294).
- Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei GW. 2021.** Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. *Communications Biology* **4**(1):228 DOI [10.1038/s42003-021-01754-6](https://doi.org/10.1038/s42003-021-01754-6).
- Wang X, Lam JY, Wong WM, Yuen CK, Cai JP, Au SW, Chan JF, To KKW, Kok KH, Yuen KY. 2020a.** Accurate diagnosis of COVID-19 by a novel immunogenic secreted SARS-CoV-2 orf8 protein. *mBio* **11**(5):e0243120 DOI [10.1128/mBio.02431-20](https://doi.org/10.1128/mBio.02431-20).
- Wang Z, Yang X, Zhou Y, Sun J, Liu X, Zhang J, Mei X, Zhong J, Zhao J, Ran P. 2020b.** COVID-19 severity correlates with weaker T-cell immunity, hypercytokinemia, and lung epithelium injury. *American Journal of Respiratory and Critical Care Medicine* **202**(4):606–610 DOI [10.1164/rccm.202005-1701LE](https://doi.org/10.1164/rccm.202005-1701LE).
- Wu S, Tian C, Liu P, Guo D, Zheng W, Huang X, Zhang Y, Liu L. 2021.** Effects of SARS-CoV-2 mutations on protein structures and intraviral protein-protein interactions. *Journal of Medical Virology* **93**(4):2132–2140 DOI [10.1002/jmv.26597](https://doi.org/10.1002/jmv.26597).
- Wu D, Wu T, Liu Q, Yang Z. 2020.** The SARS-CoV-2 outbreak: what we know. *International Journal of Infectious Diseases* **94**(395):44–48 DOI [10.1016/j.ijid.2020.03.004](https://doi.org/10.1016/j.ijid.2020.03.004).
- Young BE, Fong SW, Chan YH, Mak TM, Ang LW, Anderson DE, Lee CY, Amrun SN, Lee B, Goh YS, Su YCF, Wei WE, Kalimuddin S, Chai LYA, Pada S, Tan SY, Sun L, Parthasarathy P, Chen YYC, Barkham T, Lin RTP, Maurer-Stroh S, Leo YS, Wang LF, Renia L, Lee VJ, Smith GJD, Lye DC, Ng LFP. 2020.** Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* **396**(10251):603–611 DOI [10.1016/S0140-6736\(20\)31757-8](https://doi.org/10.1016/S0140-6736(20)31757-8).
- Yuen KS, Ye ZW, Fung SY, Chan CP, Jin DY. 2020.** SARS-CoV-2 and COVID-19: the most important research questions. *Cell and Bioscience* **10**(1):40 DOI [10.1186/s13578-020-00404-4](https://doi.org/10.1186/s13578-020-00404-4).

- Zhang Y, Chen Y, Li Y, Huang F, Luo B, Yuan Y, Xia B, Ma X, Yang T, Yu F, Liu J, Liu B, Song Z, Chen J, Yan S, Wu L, Pan T, Zhang X, Li R, Huang W, He X, Xiao F, Zhang J, Zhang H. 2021.** The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. *Proceedings of the National Academy of Sciences of the United States of America* **118(23)**:e2024202118 DOI [10.1073/pnas.2024202118](https://doi.org/10.1073/pnas.2024202118).
- Zinzula L. 2021.** Lost in deletion: the enigmatic ORF8 protein of SARS-CoV-2. *Biochemical and Biophysical Research Communications* **538(1)**:116–124 DOI [10.1016/j.bbrc.2020.10.045](https://doi.org/10.1016/j.bbrc.2020.10.045).