



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Chemico-Biological Interactions

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

Differential mutation profile of SARS-CoV-2 proteins across deceased and asymptomatic patients

Rezwanuzzaman Laskar, Safdar Ali*

Clinical and Applied Genomics (CAG) Laboratory, Department of Biological Sciences, Aliah University, Kolkata, India

ARTICLE INFO

Keywords:

SARS-CoV-2
Deceased
Asymptomatic
Deleterious
Neutral
Stability

ABSTRACT

Background: The SARS-CoV-2 infection has spread at an alarming rate with many places showing multiple peaks in incidence. Present study analyzes a total of 332 SARS-CoV-2 genome sequences from 114 asymptomatic and 218 deceased patients from twenty-one different countries to assess the mutation profile therein in order to establish the correlation between the clinical status and the observed mutations.

Methods: The mining of mutations was carried out using the GISAID CoVSurfer (www.gisaid.org/epiflu-applications/covsurfer-mutations-app) with the reference sequence 'hCoV-19/Wuhan/WIV04/2019' present in NCBI with Accession number NC-045512.2. The impact of the mutations on SARS-CoV-2 proteins mutation was predicted using PredictSNP1 (loschmidt.chemi.muni.cz/predictsnp1) which is a meta-server integrating six predictor tools: SIFT, PhD-SNP, PolyPhen-1, PolyPhen-2, MAPP and SNAP. The iStable integrated server (predictor.nchu.edu.tw/iStable) was used to predict shifts in the protein stability due to mutations.

Results: A total of 372 variants were observed in the 332 SARS-CoV-2 sequences with several variants present in multiple patients accounting for a total of 1596 incidences. Asymptomatic and deceased specific mutants constituted 32% and 62% of the repertoire respectively indicating their partial exclusivity. However, the most prevalent mutations were those present in both. Though some parts of the genome are more variable than others but there was clear difference between incidence and prevalence. Non-structural protein 3 (NSP3) with 68 variants had a total of only 105 incidences whereas Spike protein had 346 incidences with just 66 variants. Amongst the Deleterious variants, NSP3 had the highest incidence of 25 followed by NSP2 (16), ORF3a (14) and N (14). Spike protein had just 7 Deleterious variants out of 66.

Conclusion: Deceased patients have more Deleterious than Neutral variants as compared to the asymptomatic ones. Further, it appears that the Deleterious variants which decrease protein stability are more significant in pathogenicity of SARS-CoV-2.

1. Introduction

The causative agent of ongoing COVID-19 global pandemic is Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) which belongs to family *Coronaviridae* characterized by single strand positive sense RNA genome [1]. As of 31st March 2021, there were 5,52,566 active cases; 1,14,34,301 discharged cases and 1,62,468 deaths in India due to SARS-CoV-2 (<https://www.mohfw.gov.in/>). The same time, as per WHO there were 2,73,49,248 confirmed cases of COVID-19, including 27,87,593 deaths worldwide due to COVID-19 (covid19.who.int).

As the SARS-CoV-2 infection spread at an alarming rate and many places showed multiple peaks in incidence [2–4], the virus accrued

mutations in the process. There have been multiple studies to assess and understand the impact of these mutations [5–8]. These reports have focused on novel and recurrent variations and their impact on infectivity and antigenicity. However, the assessment of impact of mutations on SARS-CoV-2 is an ongoing process. Since the mutational profile of SARS-CoV-2 has been reported to be country specific, the analysis of mutations in a country wise manner assumes significance. For example, mutations at nucleotide positions 17746, 17857 and 18060 are exclusively present in North America [9]. Further, some proteins of SARS-CoV-2 play a critical role in viral infection and pathogenesis. For instance, the Spike (S) protein whose interaction with the angiotensin converting-enzyme 2 (ACE2) receptor is pivotal for viral entry into host cell. There are 17 residues in the receptor binding domain (RBD) of S

* Corresponding author. Department of Biological Sciences, Aliah University, IIA/27, Newtown, Kolkata, 700160, India.

E-mail addresses: rezwanuzzaman.laskar@gmail.com (R. Laskar), ali@aliah.ac.in, ali@aliah.ac.in (S. Ali).

<https://doi.org/10.1016/j.cbi.2021.109598>

Received 13 April 2021; Received in revised form 8 July 2021; Accepted 21 July 2021

Available online 23 July 2021

0009-2797/© 2021 Elsevier B.V. All rights reserved.

Table 1
Summary of proteins of SARS-CoV-2 and mutations observed in the study.

S No	Protein Feature	ORF Name	Protein Name	RefSeq protein Id	AA Length	Function	Observed Mutations	Mutations/ 100 Residues
1	Non-Structural Protein	ORF1a	NSP3 (Papain like Proteinase/PLPro)	YP_009725299	1945	Proteolytic cleavage [31].	68	3.5
2	Structural Protein	ORF 2	Surface Glycoprotein (S Protein)	YP_009724390	1273	Crucial for viral infection, integration, and ingress into the host cell [32].	66	5.18
3	Structural Protein	ORF9	Nucleocapsid Phosphoprotein (N Protein)	YP_009724397	419	Binds with RNA and facilitates the assembly of vRNPs (viral RNA-protein) into virions [33].	29	6.92
4	Non-Structural Protein	ORF1a	NSP2	YP_009725298	638	Disruption of the infected cells microenvironment [31].	27	4.23
5	Non-Structural Protein	ORF1b	NSP12 (RNA Dependent RNA Polymerase)	YP_009725307	932	This RdRp (RNA-dependent RNA-polymerase) protein aids in viral replication and transcription [31].	21	2.25
6	Accessory Protein	ORF3a	ORF3a Protein	YP_009724391	275	Viroporin (ion channel protein) that facilitates viral release [26].	20	7.27
7	Non-Structural Protein	ORF1b	NSP13 (Helicase)	YP_009725308	601	Unwinding RNA helices [34]	18	3
8	Non-Structural Protein	ORF1b	NSP14 (Exo-ribonuclease)	YP_009725309	527	Proofreading exoribonuclease [35].	17	3.23
9	Non-Structural Protein	ORF1a	NSP4 (Contains Transmembrane Domain 2)	YP_009725300	500	Involved in the creation of two-membrane vesicles [36].	14	2.8
10	Non-Structural Protein	ORF1b	NSP15 (Endo-RNase)	YP_009725310	346	Destroys viral dsRNA in order to elude recognition by the host [37].	14	4.05
11	Non-Structural Protein	ORF1a	NSP5 (3C-like proteinase/3CLPro)	YP_009725301	306	Cleaving polypeptides to form nonstructural proteins [31].	11	3.59
12	Non-Structural Protein	ORF1a	NSP6 (Putative Transmembrane Domain)	YP_009725302	290	Inhibits auto-phagosomal expansion and the degradation of viral components in the lysosome [36].	9	3.1
13	Accessory Protein	ORF8	ORF8 Protein	YP_009724396	121	Inhibits IFN-I signaling [35].	9	7.44
14	Non-Structural Protein	ORF1b	NSP16 (2'-O-Ribose-Methyltransferase)	YP_009725311	298	Pivotal role in mRNA cap methylation for host immune system evasion [31].	8	2.68
15	Accessory Protein	ORF7a	ORF7a Protein	YP_009724395	121	Induces IFN-I antagonism and inhibit STAT1 phosphorylation [27].	8	6.61
16	Structural Protein	ORF5	Membrane Glycoprotein (M Protein)	YP_009724393	222	M protein interacts with E, S, and N proteins to play a vital role in viral assembly and budding [38].	7	3.15
17	Structural Protein	ORF4	Envelope Protein (E Protein)	YP_009724392	75	Viroporin involved in viral synthesis, assembly, and release [39].	6	8
18	Non-Structural Protein	ORF1a	NSP1 (Leader Protein)	YP_009725297	180	Inhibitory effect on host gene expression and destroys host mRNA [28].	5	2.78
19	Accessory Protein	ORF6	ORF6 -Protein	YP_009724394	61	Involved in mitigating the host's immune reaction and viral replication [40].	5	8.2
20	Non-Structural Protein	ORF1a	NSP8 (Primase)	YP_009725304	198	Interacts with NSP7 to form a hexamer and functions as a primase during replication [41].	4	2.02
21	Non-Structural Protein	ORF1a	NSP10	YP_009725306	139	Stimulating the activities of NSP14 and NSP16 [36].	3	2.16
22	Non-Structural Protein	ORF1a	NSP7 (Primase)	YP_009725303	83	Interacts with NSP8 to form a hexamer and functions as a primase during replication [42].	2	2.41
23	Non-Structural Protein	ORF1a	NSP9 (RNA-Binding Protein)	YP_009725305	113	Serves as a single stranded RNA binding protein in viral replication [28]	1	0.88
24	Non-Structural Protein	ORF1a	NSP11	YP_009725312	13	Its functionality is undefined	0	0
25	Accessory Protein	ORF7b	ORF7b Protein	YP_009725318	43	Induces IFN-I antagonism and inhibits STAT1 & STAT2 phosphorylation [35]	0	0
26	Accessory Protein	ORF10	ORF10 Protein	YP_009725255	38	Its functionality is undefined	0	0

Table 2
Country-wise distribution of samples used in the study.

S No	Country	Total	Asymptomatic	Deceased
01	Bangladesh	5	5	0
02	Belgium	2	1	1
03	Brazil	62	2	60
04	Colombia	3	0	3
05	Costa Rica	2	0	2
06	Czech Republic	3	1	2
07	Dominican Republic	3	3	0
08	India	68	28	40
09	Indonesia	2	0	2
10	Italy	3	1	2
11	Japan	59	59	0
12	Kuwait	2	2	0
13	Lebanon	1	0	1
14	Mexico	4	0	4
14	Oman	1	0	1
16	Russia	3	0	3
17	Saudi Arabia	70	0	70
18	South Africa	1	0	1
19	Sri Lanka	1	0	1
20	Turkey	8	8	0
21	United States	29	4	25
		332	114	218

protein which makes contact with the ACE2 receptor and mutations affecting this interaction will directly impact viral pathogenesis [10,11]. A brief summary of all the proteins of SARS-CoV-2 and their function has been provided in Table 1. The proteins have been listed in descending order of number of mutations observed in present study.

In a previous study, we performed the mutational analysis of 611 genomes from India extracted in June 2020 and analyzed its impact on proteins. Therein we observed a difference in mutation profile in viral genomes across deceased and asymptomatic patients. While only 11% disease mutations were present in genomes from asymptomatic people,

the corresponding value from deceased patients was more than three folds at 38% implying a possible correlation between the nature of mutations and clinical outcome. However, one lacuna of the study was very limited number of samples with available clinical metadata (30 asymptomatic and 15 deceased) [12]. Presently, we expand our sample size by using sequences from twenty-one different countries with an attempt to ascertain the differential mutation profile of deceased and asymptomatic samples in order to understand the possible clinical correlation, if any.

2. Methods

2.1. Sequence Congregation

On 10th November 2020, we retrieved 6853 sequences from GISAID (www.gisaid.org) using the data filter ~ virus name: hCoV-19 - Host: Human - Complete – High Coverage. Subsequently, they were screened for availability of clinical status which revealed 122 and 542 sequences from Asymptomatic and Deceased patients respectively. The final screening parameter applied was that of age wherein all samples of over 60 years were excluded. There are multiple risk factors associated with mortality due to SARS-CoV-2 which include age, hypertension and diabetes. Since a detailed clinical profile was not available for most samples only age was used as an exclusion criterion. Multiple reports suggest deaths of old age people infected with SARS-CoV-2 is often due to co-morbidities and a weakened immune system [13,14]. Since our aim was to ascertain the impact of mutations, if any, on the deaths due to SARS-CoV-2, so we excluded the samples where the chance of other factors contributing to death is high.

Thereon, a total of 332 SARS-CoV-2 sequences from 114 Asymptomatic and 218 Deceased patients were included in the study from twenty-one different countries. Details of the studied sequences have been provided in Table 2 and Supplementary file 1. These included 200

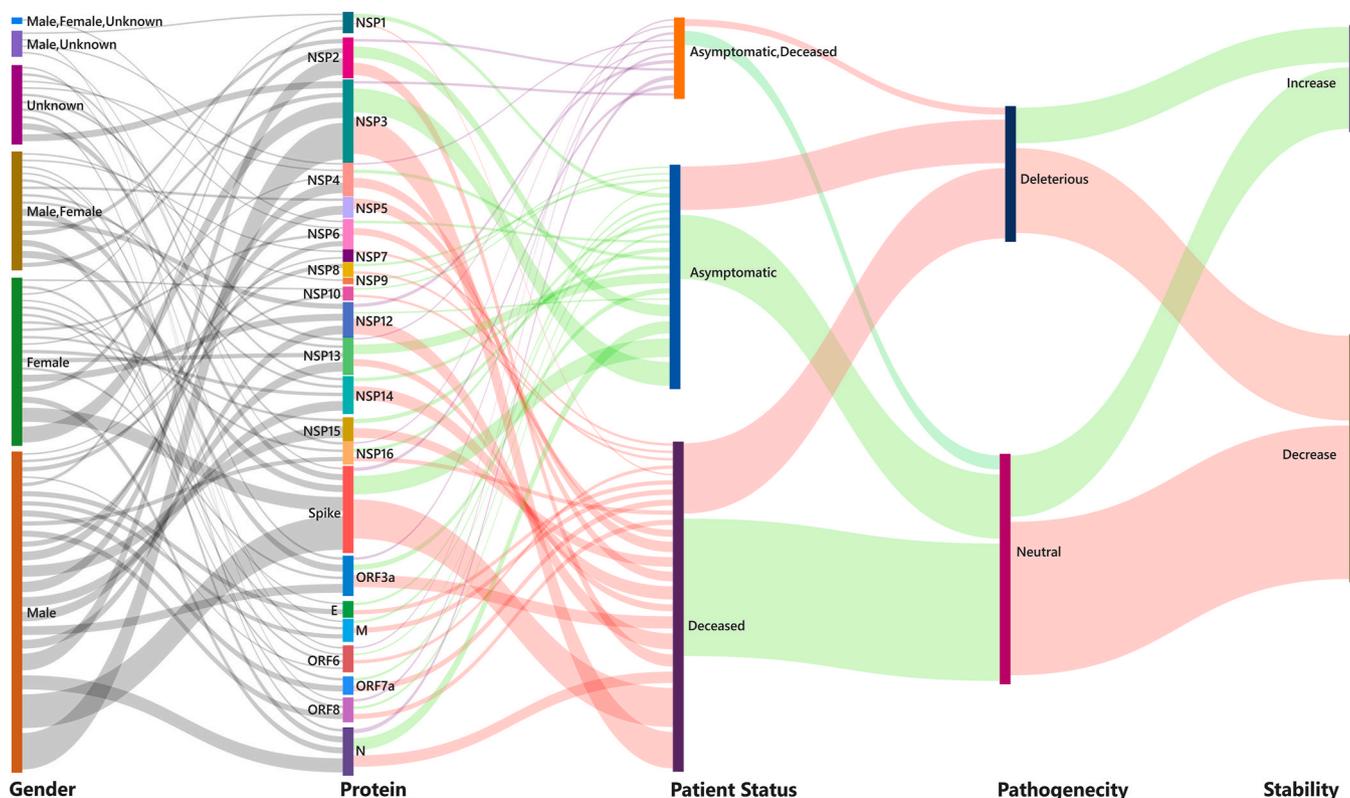


Fig. 1. Distribution of variant sites of SARS-CoV-2 proteins across Gender; Patient status (Deceased/Asymptomatic); Pathogenicity (Deleterious/Neutral) and Stability.

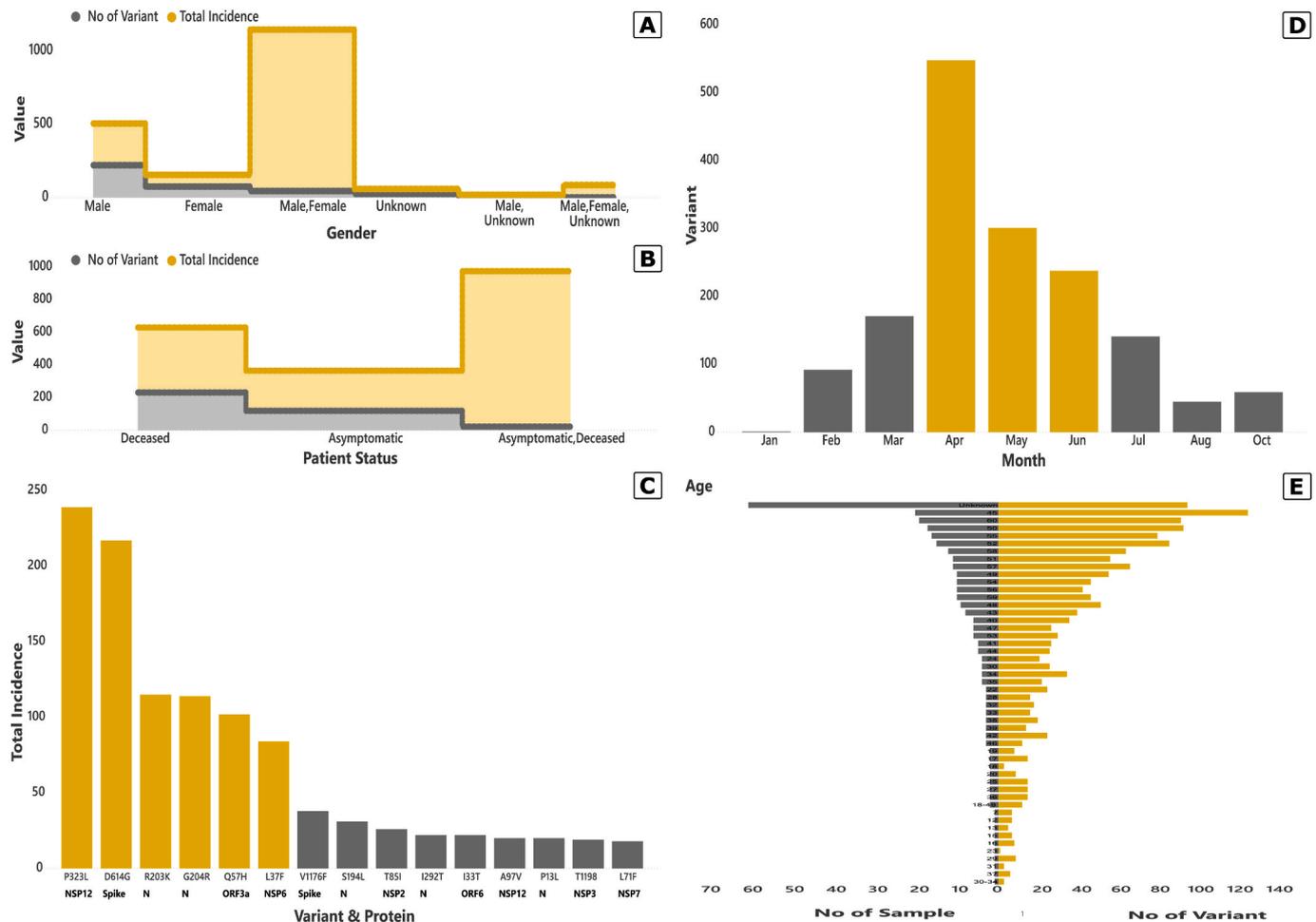


Fig. 2. Details of mutations of SARS-CoV-2 proteins. A) Gender wise distribution of variants and its prevalence in studied genomes. B) Distribution of observed variants according to clinical status of patients as per Deceased/Asymptomatic/Both C) Most prevalent variants across proteins observed in the study. D) Timeline of incidence of observed variations. E) Age wise distribution of samples of the present study.

males, 72 females and 60 patients for which gender information was not available. There were ten countries which had no asymptomatic patients and five countries with no deceased patients. It is quite plausible that the absence of sequences from the said countries can be attributed to lack of metadata at GISAID repository and do not necessarily imply the absence of such patients in the countries.

2.2. Mining of mutations in proteins

The mining of non-synonymous mutations from the selected 332 sequences was done utilizing the GISAID CoVServer (www.gisaid.org/epiflu-applications/covserver-mutations-app) web resources for sequence variance analysis. Sequence datasets were aligned with the reference sequence 'hCoV-19/Wuhan/WIV04/2019'. The reference amino acid sequence of each protein for further analysis was downloaded from the same server. The reference sequence is identical to the NCBI sequence with Accession number NC_045512.2 from Wuhan, China. It has also been used as reference for our earlier studies on SARS-CoV-2 genomes from India [12,15].

2.3. Pathogenicity Prediction of mutations

The impact of the mutations on SARS-CoV-2 proteins mutation was estimated using PredictSNP1 (loschmidt.chemi.muni.cz/predictsnp1) which is a meta-server integrating six predictor tools: SIFT, PhD-SNP,

PolyPhen-1, PolyPhen-2, MAPP and SNAP. The SNAP and PhD-SNP tools are based on supervised machine learning algorithms that have been trained in massive datasets to "learn" to distinguish between pathogenic and benign variants. The protein sequence and structure method predict how mutations influence the protein phenotype on the basis of the SNP position in the protein structure behind the PolyPhen-1 and PolyPhen-2 predictors. The MAPP and SIFT tools, on the other hand, determine the pathogenicity of mutations based on the conservation of specific amino acids by sequence and evolutionary conservation methods across various species. PredictSNP extracted the individual score of these tools to homogenize the assessment and generate its own confidence score as a percentage of expected accuracy, ranging from 0 to 100%. It has also designed three distinct datasets like PMD, MMP and OVERFIT in order to eliminate bias, duplicity and inconsistency [16]. Subsequently, the variants are categorized as 'Neutral' or 'Deleterious', and these predictions are more robust and accurate than the prediction provided by any individual predictor.

2.4. Stability shifts of protein mutations

The iStable integrated server (predictor.nchu.edu.tw/iStable) encompasses two machine-learning based predictor iMutant and MUpro along with thermodynamics parameters in order to predict shifts in the protein stability related to mutations. i-Mutant, an SVM-based tool, predicts protein stability attributable to mutant form in free energy

Table 3

Distribution of Deleterious (Red) and Neutral (Black) mutations along with D/N ratio (Deleterious/Neutral) across different proteins of SARS-CoV-2 in Deceased (De) and Asymptomatic (As) patients. Both (Bo) denotes presence in Asymptomatic as well as Deceased patients.

S No	Protein	Total Variations (De+As+Bo)	Deceased (De)	Asymptomatic (As)	Both (Bo)	Deceased (De)	D/N Ratio	Asymptomatic (As)	D/N Ratio	Both (Bo)	D/N Ratio
1	E	6(4+2+0)	C43F L21F; T30I; P71S	R38C I46V	Nil	1, 3	0.33	1, 1	1	0, 0	NA
2	M	7(5+2+0)	A2V; D3G; L29F; V66L; H125Y	A98S; C33F	Nil	0, 5	0	0, 2	0	0, 0	NA
3	N	29(12+13+4)	A119S; S193I; S194L; T247I; K373N; D402Y; Q418H R41L; L139F; H145Y; S327L; K374N	D225Y; Q9H; R195I; S197L; S202N Q7K; D103N; I351V; P199S; H300Y; T334A; T379I; A397V	G204R; P13L R203K; I292T	7, 5	1.4	5, 8	0.625	2, 2	1
4	ORF3a	20(13+5+2)	S40L; L41H; A51S; I62M; A110V; R126S; G172V; Q185K; G251V; S253P I118T; M125I; V225F	T89I; G196V; G224C L15F; L52I	Q57H S171L	10, 3	3.33	3, 2	1.5	1, 1	1
5	ORF6	5(3+1+1)	I11R; R20S; W27L	V5F	I33T	3, 0	∞	1, 0	∞	1, 0	∞
6	ORF7a	8(6+2+0)	V74F; S83L; V93F; Q94E A8S; V108L	T39I A105V	Nil	4, 2	2	1, 1	1	0, 0	NA
7	ORF8	9(5+2+2)	Q23H D34G; A51S; A51V; A65D	D119G; V99L	L84S; S24L	1, 4	0.25	0, 2	0	0, 2	0
8	NSP1	5(1+4+0)	R124C	R24C; D75N; I114T	Nil	0, 1	0	3, 1	3	0, 0	NA
9	NSP2	27(13+12+2)	D23Y; E66G; P181T; S211F; T371I; Q496P I21V; P129L; S196L; A247V; R380C; T429I; T434I	E57K R27C; K110N; M141T; A159V; G339S; K347N; Q427H; L520F; V577F I120F; S430L; V198I;	T85I A318V	6, 7	0.86	9, 3	3	1, 1	1
10	NSP3	68(40+26+2)	D112Y; D178Y; A256V; G337C; L425P; T428I; P466L; L623I; L689F; R810C; L862F; K1083I; D1121G; L1244F; G1476S; A1914D A1S; S126L; P200S; G212S; P389L; V481L; K487N; A496V; T504P; S716I; T724I; P822L; K1077N; K1211N; P1228L; S1314G; A1324V; S1424P; A1431V; S1670F; S1682F; K1693N; V1795I; A1872V	S1197R; D163Y; F430S; G470S; G470V; T779I; K945N; V1768G; S1807F P153S; I341M; K379E; K412N; I458T; A480V; A488T; M494I; P723S; A994D; K1042N; T1072I; C1114I; P1132S; M1278K; D1674N; V1762A	A861V; T1198K	16, 24	0.67	9, 17	0.53	0, 2	0
11	NSP4	14(10+3+1)	S148P; G309C W6C; T204I; T214I; A231V; T295I; M366I;	G196D; F308Y M33I	A307V	2, 8	0.25	2, 1	2	0, 1	0

			A457V; S59P								
12	NSP5	11(11+0+0)	A266V; L89F; T196M; V18I K12R; G15S; T45I; G71S; P96S; V125I; T198I	Nil	Nil	4, 7	0.57	0, 0	NA	0, 0	NA
13	NSP6	9(7+2+0)	V190F; I162T A88V; L125F; V149F; Y153F; M183I	C113F; L37F	Nil	2, 5	0.4	0, 2	0	0, 0	NA
14	NSP7	2(2+0+0)	L71F; S25L	Nil	Nil	2, 0	∞	0, 0	NA	0, 0	NA
15	NSP8	4(2+2+0)	A74V T123I	G144S; V26I	Nil	1, 1	1	0, 2	0	0, 0	NA
16	NSP9	1(0+1+0)	Nil	L42F	Nil	0, 0	NA	0, 1	0	0, 0	NA
17	NSP10	3(2+1+0)	K124R; T111I	T101I	Nil	0, 2	0	0, 1	0	0, 0	NA
18	NSP12	21(17+1+3)	D140G; T248I; L308M; V354A; V354G; S814P L8F; N9K; M196I; P227L; A250V; T252N; L329I; N491S; V880I; A878S; H928Q	V354L	A97V; L638F P323L	6, 11	0.55	1, 0	∞	2, 1	2
19	NSP13	18(7+10+1)	S74P; T115I L297F; D466N; T501I; I258V; I334V	P47S; A237T; S485L; P504L; Y541C; M576I G54S; S236I; K524N; A553V	S166A	2, 5	0.4	6, 4	1.5	0, 1	∞
20	NSP14	17(14+3+0)	S28I; R163C; L177F; T215I N3H; K32R; P140S; I45K; V161I; C216F; A320V; K349N;	D345G; H26Y; T206I	Nil	4, 10	0.4	0, 3	0	0, 0	NA
21	NSP15	14(10+4+0)	Q354H; T372I S287L; Y225F T33I; V38F; A171V; G129S; E223A; K259N; E260K; P270T;	I143N; R206S V172L; S261L	Nil	2, 8	0.25	2, 2	1	0, 0	NA
22	NSP16	8(4+3+1)	T117I; I267T A116V; K160R	K263N M17I; T35I	S166A	2, 2	1	1, 2	0.5	0, 1	0
23	Spike	66(43+20+3)	C136G; Y489H; S939F P9L; S12F; T22I; A27S; L54F; F59S; A67V; R78M; Y144F; M153I; H146Q; Q183H; V213A; A264S; T299I; V382L; T478K; E484K; S494P; N501Y; T572I; D574Y; Q613H; A647S; A688S; A688V; T791I; P812L; A879S; N978D; L1063F; T1117I; G1124V; D1153Y; P1162S; V1176F; N1192K; I1225T; M1237L; I870V	S46L; D88H; G769V; D1084Y D627P; G181V; G476S; I119V; K1086R; K77M; K854N; M1229I; N679K; P1079T; P681R; Q218H; R765L; T29I; V1228L; V47I	D614G; E583D; L5F	3, 40	0.075	4, 16	0.25	0, 3	0



Fig. 3. Country wise distribution of variants of SARS-CoV-2 proteins.

change value [DDG = DG (New Protein) – DG (Wild Type) in Kcal/mol]. It classifies the variant in terms of decreasing (DDG<0) or increasing (DDG>0) protein stability. MUpro, an SVM and neural network-based tool, evaluates the mutational impact of protein stability on predictive confidence score (Conf. Score). The score varies from –1 to 1, with Conf. Score <0 signifies a decrease in stability and Conf. Score >0 as an increase in protein stability. iStable utilizes sequence information and predicts the meta-result as an increase or decrease in stability in terms of confident score, where a higher score indicates more confidence in the prediction [17].

3. Results and discussion

3.1. Mutation incidence and prevalence

A total of 372 variants were observed in 332 SARS-CoV-2 sequences with several variants being incident in multiple patients accounting for a total of 1596 incidences. The distribution of variants across different proteins of SARS-CoV-2 have been shown in Fig. 1, Table 1 and Supplementary file 2. The distribution and incidence of variants can be analyzed through several aspects.

First, the incidence of mutations with respect to gender as summarized in Fig. 2A. The patients for whom gender information was not available have been mentioned as unknown. There were 45 mutations (12%) present in both males and females and these variants were the most prevalent ones as they accounted for around 69% of the total incidences. Also, the male and female specific 221 and 75 mutations (59% and 20%) accounted for a meagre 17% and 5% of the total incidence respectively. Partly this difference can be attributed to the skewed sample set in favor of the males. Klein and Flanagan have reviewed the studies assessing the variable nature of immune responses in males and

females [18] but since the variants most incident are present in both genders in our study and assuming the variations can potentially alter the immune response, the chances of it doing so in a gender-dependent manner seems rare.

Though there were 60 patients with unknown gender but the fact that 90% of the overall mutations in the studied population is present in males, females or both means that the data from samples with unknown gender will not have much of an impact on gender wise distribution and incidence of mutants. Further, almost all of these 60 patients were asymptomatic and belong to the Diamond Princess cruise ship from Japan which has become a benchmark study about how wearing masks reduces viral load wherein even if the person gets infected, chances of it being mild or asymptomatic are enhanced substantially [19,20].

Subsequently, we moved to the second aspect of our analysis in terms of patient status where we screened for mutations specific to the deceased or asymptomatic patients. The distribution and incidence of mutations as per patient status has been shown in Fig. 2B and Supplementary file 2. Interestingly, 22 mutations present in both deceased and asymptomatic patients accounted for 952 (60%) of the total incidence of mutations. Asymptomatic specific mutants encompassing 119 variants (32%) accounted for only 246 (15%) of the total observed mutations whereas the corresponding values for Deceased specific mutants was 231 (62%) variants with 398 (24%) total incidences respectively. The details of these mutations have been provided in Table 3 and Supplementary file 2 and their impact on proteins is discussed later.

The impact of codon bias in SARS-COV-2 genomes on incidence of mutations has been reported wherein the most abundant mutations were C > U (46%), G > U (18.2%), U > C (9.4%) and A > G (8.8%) implying an increase in U as a viral strategy to adapt to host codon usage. This results in some mutations being more prevalent than others [21]. Evidently, in the present study as well, some mutations are getting

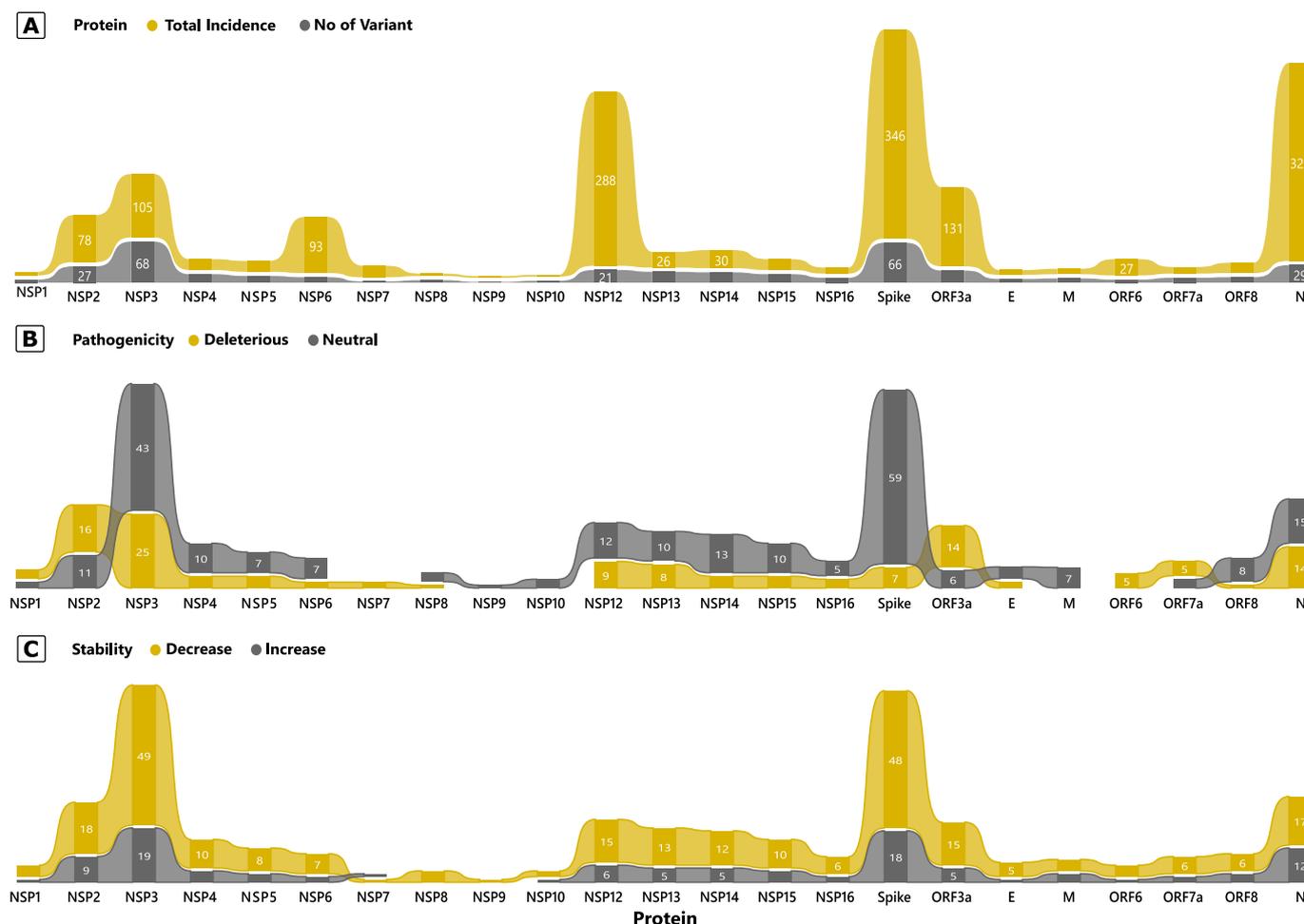


Fig. 4. Mutation incidence and its impact on SARS-CoV-2 proteins. **A)** Number of variants and its prevalence in studied genomes. **B)** Pathogenicity Prediction of the variants in terms of Deleterious or Neutral. **C)** Stability shift prediction of the observed variants.

represented at a much higher rate as compared to others. The fifteen most prevalent mutations along with their incident frequencies and localizations have been shown in Fig. 2C. P323L (NSP12) is the most prevalent variant present followed by D614G (Spike protein) with over two hundred incidences each. Most of these prevalent variants are common to both males and females, which is expected as per gender wise distribution data.

The timeline for accumulation of mutants has been shown in Fig. 2D and evidently the maximum number of mutations accrued in April 2020 wherein the incidence of COVID-19 was at its peak. The age wise distribution of mutants has been shown in Fig. 2E. An absolute correlation between age and incidence cannot be drawn due to uneven representation of all age groups. Furthermore, since the samples are from different countries and the virus is behaving differently across geographical locations, the data has to be interpreted with caution. Another study encompassing 48,635 SARS-CoV-2 sequences reported a total of 353,341 mutation events compared to Wuhan reference genome. Of these, 256 samples were identical to the reference whereas 48,379 samples possessed at least one mutation. An average of 7.23 mutations per sample was reported in the study which also highlighted country specific average incidence of mutations [22]. Thereby, looking at the mutation incidence with reference to individual countries seemed rational.

We analyzed the data in terms of number of mutants per sample for each country as represented in Fig. 3 and Supplementary file 4. Saudi Arabia with maximal representation of 70 samples in the study was contributing 3.93 variants per sample which is amongst the lowest in the

group. Contrastingly, Bangladesh with highest value of 10 variants per sample is being represented by just 5 samples in the study. Thus, it is explicit that some nations are exhibiting more variations in SARS-CoV-2 as compared to others. Differential health and gene profile of individuals might be factors contributing to it. We also looked at the variant per sample data for deceased and asymptomatic samples respectively for each country as shown in Supplementary file 4. Bangladesh had the highest variants per sample of 10 followed by India (6.36) for asymptomatic patients whereas it was the highest of 9.5 for Indonesia followed by Brazil (6.75) for deceased patients. The fact that some countries had no representation in asymptomatic or deceased samples makes any direct inference plausible but a geographical dependent evolution of the virus is surely supported by the observed data.

3.2. Impact of mutations on proteins of SARS-CoV-2

The impact of mutations on SARS-CoV-2 proteins was assessed using three aspects as highlighted in Fig. 4. These include distribution of variants across SARS-CoV-2 proteins; pathogenicity of the variants in term of deleterious or neutral and assessing the stability of proteins in presence of variants.

The distribution of variants across different SARS-CoV-2 proteins revealed some interesting observations. First, incidence of mutations in some proteins was relatively higher than others as evident in Table 1, Fig. 4a and Supplementary file 2. NSP3 housed a maximum of 68 variants closely followed by Spike protein with 66 variants. N protein and NSP2 with 29 and 27 variants respectively were a distant third and

Table 4

Correlation between Deleterious mutations from Deceased individuals and Protein Stability. (Variants with increased stability have been shown in green).

Selection of Variants	S No	Protein	Protein Feature	Variant Name	No of variant	Decrease Stability (DS)	Increase Stability (IS)
Deceased-Deleterious-Decrease Stability De-D-DS	1	E	Structural Protein	C43F	1	1	0
	2	ORF7a	Accessory Protein	Q94E, S83L, V74F, V93F	4	4	0
	3	ORF8	Accessory Protein	Q23H	1	1	0
	4	NSP6	Non-Structural Protein	I162T, V190F	2	2	0
	5	NSP8	Non-Structural Protein	A74V	1	1	0
	6	NSP12	Non-Structural Protein	D140G, L308M, S814P, T248I, V354A, V354G	6	6	0
	7	NSP13	Non-Structural Protein	S74P, T115I	2	2	0
	8	NSP16	Non-Structural Protein	I267T, T117I	2	2	0
	9	S	Structural Protein	C136G, S939F, Y489H	3	3	0
Deceased-Deleterious-Decrease/Increase Stability De-D-DS/IS	1	NSP4	Non-Structural Protein	G309C, S148P	2	1	1
	2	ORF6	Accessory Protein	I11R, R20S, W27L	3	2	1
	3	NSP5	Non-Structural Protein	A266V, L89F, T196M, V18I	4	3	1
	4	NSP7	Non-Structural Protein	L71F, S25L	2	1	1
	5	NSP15	Non-Structural Protein	S287L, Y225F	2	1	1
	6	NSP14	Non-Structural Protein	L177F, R163C, S28I, T215I	4	2	2
	7	NSP2	Non-Structural Protein	D23Y, E66G, P181T, Q496P, S211F, T371I	6	3	3
	8	N	Structural Protein	A119S, D402Y, K373N, Q418H, S193I, S194L, T247I	7	3	4
	9	ORF3a	Accessory Protein	A110V, A51S, G172V, G251V, I62M, L41H, Q185K, R126S, S253P, S40L	10	6	4
	10	NSP3	Non-Structural Protein	A1914D, A256V, D1121G, D112Y, D178Y, G1476S, G337C, K1083I, L1244F, L425P, L623I, L689F, L862F, P466L, R810C, T428I	16	11	5

fourth. However, a greater number of mutations does not necessarily imply it being prevalent in the studied samples. To ascertain the same, we calculated the incident mutations per hundred amino acids for all the proteins as shown in Table 1. The proteins have been sorted as per number of observed mutations. NSP3 with 68 mutations had 3.5 mutations per hundred residues whereas ORF8 with just 9 mutations had a

corresponding value of 7.44 indicating an unequal distribution of mutations across SARS-COV-2 proteins. A bias in incidence of mutations in SARS-COV-2 proteins has been previously reported as well with surface glycoprotein, nucleocapsid, ORF1ab, and ORF8 exhibiting a higher frequency of mutations as compared to envelop, membrane, ORF6, ORF7a and ORF7b. Further, emergence and stability of variants

accounted for their widespread global distribution [23,24]. Thus, some parts of the genome are more variable whereas others are more conserved. These have led to emergence of novel variants around the world which have had varied clinical manifestations. Secondly, there was a clear difference between incidence and prevalence. NSP3 with 68 variants had total incidence of only 105 whereas Spike protein had 346 incidences (66 variants). Comparatively, N protein and NSP12 had 328 (29 variants) and 288 (21 variants) incidences in the study. This clearly implies that a greater number of mutations does not necessarily imply higher prevalence.

In order to ascertain the significance of differential incidence and prevalence, we analyzed the pathogenicity of the variants as Deleterious or Neutral which has been shown in Fig. 4b, Table 3 and Supplementary file 3. Their presence across Asymptomatic and Deceased samples across different countries has been discussed in the next section. The basic premise for the study was that a protein having more variants will be contributing to the viral evolution only if they are Deleterious. Neutral mutations wouldn't be affecting the protein per se. In terms of incidence of Deleterious variants, NSP3 had the highest of 25 followed by NSP2 (16), ORF3a (14) and N (14). The NSP3 protein is responsible for proteolytic activity but is not the only such protein. NSP5 also has similar activity and has 4 Deleterious mutations associated exclusively to deceased patients. Spike protein had just 7 Deleterious variants out of 66 which partly explains that despite of so many variants it has not much impacted the viral pathogenesis yet. Though several mutations like K417 N, E484K, N501Y, D614G and P681H in the S protein along with the importance of RBD domain to the virulence of SARS-CoV-2 have been reported but with emerging variants their impact needs to be constantly monitored [10,11,25].

Interestingly, only four proteins NSP1, NSP2, ORF3a and ORF7a had more Deleterious than Neutral mutants. NSP1 is involved in inhibiting host gene expression and NSP2 in disrupting the microenvironment for infected cells. ORF3a and ORF7a are associated with virion release and evading host immune response respectively. The presence of more deleterious mutations therein suggests viral adaptability [26–28]. Further, NSP7 and ORF6 had only Deleterious mutants indicating their susceptibility to mutations (Fig. 4b, Table 3). Thus, we can say that even a protein with very few mutations can be a pivotal factor in viral evolution.

Assessment of the protein stability in lieu of mutations was thereby ascertained and has been represented in Fig. 4c. Majority of variants resulted in a decrease in protein stability across all proteins. The three proteins which had maximum variants which increased protein stability included NSP3 (19), Spike (18) and N (12). Also, NSP3 (49) and Spike (48) had highest number of variants which decreased protein stability. The impact of these variants when discussed individually is significant but their presence in population is not in isolation and hence the cumulative impact of mutations incident together is also required.

3.3. Differential mutational profile across asymptomatic and deceased samples

The distribution of Deleterious and Neutral mutations across asymptomatic and deceased patients has been shown in Table 3 and Supplementary file 3. There are 12 proteins in which no variant was present in both asymptomatic and deceased samples suggesting a correlation between mutational status and disease profile. Contrastingly, N protein had the highest of 4 variants across both sample sets. Moreover, except for N protein, NSP1, NSP3, NSP9 and NSP14, all other proteins had more variants associated with deceased patients as compared to asymptomatic ones. NSP5 and NSP7 had no mutations from asymptomatic patients whereas only NSP9 had no variants coming from deceased patients.

Subsequently we analyzed the ratio of Deleterious to Neutral mutations across asymptomatic and deceased samples as shown in Table 3. Most proteins had a higher D/N ratio for deceased samples indicating

the implication of Deleterious mutations therein. The highest D/N ratio was 3.33 for ORF3a followed by 1.4 for N protein. The fact that deceased patients have more Deleterious than Neutral variants as compared to the symptomatic ones is suggestive of these mutations being correlated to the disease status of the samples. Therefore, the exclusive mutations for deceased and asymptomatic samples can serve as a benchmark for management of patients.

Transversion (11083G > T) in ORF1ab gene leading to substitution of leucine to phenylalanine in NSP6 has been shown to differentiate between symptomatic and asymptomatic patients. Further, two mutations (26,144G > T) and (1,397G > A) were reported to distinguish between symptomatic and asymptomatic patients respectively [29]. It has also been shown that asymptomatic patients harbor mutations at 9 nucleotide positions (C2939T; C3828T; G21784T; T21846C; T24631C; G28881A; G28882A; G28883C; G29810T) associated with several non-synonymous substitutions across ORF1ab (P892S; S1188L), S (K74 N; I95T) and N (R203K, G204R) proteins [30]. Present study also exhibited the two variants of N protein. However, they are present in both males and females of asymptomatic as well as deceased patients. This signifies the constantly changing genetic landscape in SARS-CoV-2.

To further construe the link between the deceased host and the non-synonymous mutations, we analyzed the mutated protein clusters as follows: From Deceased Individuals (De) with Deleterious (D) pathogenicity and Increase in Stability (IS) and From Deceased Individuals (De) with Neutral (N) pathogenicity and Decrease in Stability (DS). The details have been shown in Table 4 and Supplementary file 3. When we analyzed mutations only from deceased patients and with deleterious pathogenicity, we observed that there were nine proteins in which all such mutations had decreasing stability suggesting the possibility of the decreased stability contributing to enhanced pathogenicity. Moreover, there were ten proteins which had mutations with increasing as well as decreasing stability. Amongst these there was N protein which had four variants with increasing stability as compared to three with decreasing stability prediction. All other proteins had more stability decreasing variants as compared to increasing ones. Thus, we can say that the variants with decreasing stability are more significant in pathogenicity of SARS-CoV-2.

4. Conclusions

A total of 372 variants were observed in 332 SARS-CoV-2 sequences with several variants being incident in multiple patients accounting for a total of 1596 variants. Since, some countries had no representation in asymptomatic or deceased samples, inference about geographical correlation is not plausible from present dataset. However, the deleterious pathogenicity mutations found in deceased patients can serve as guide for patient management which can be based on nine proteins (E; ORF7a; ORF8; NSP6; NSP8; NSP12; NSP13; NSP16 and S) in which twenty-two such mutations had decreasing stability implying its contribution to enhanced pathogenicity.

Ethics approval

Not Applicable.

Availability of data and materials

All data pertaining to the study has been provided as Supplementary material of the manuscript.

CRediT authorship contribution statement

Rezwanuzzaman Laskar: Methodology, Investigation, Formal analysis, Validation. **Safdar Ali:** Conceptualization, Supervision, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the Department of Biological Sciences, Aliah University, Kolkata, India for all the financial and infrastructural support provided. Authors acknowledge all the authors associated with originating and submitting laboratories of the sequences from GISAID's EpiFlu™ (www.gisaid.org) Database on which this research is based.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cbi.2021.109598>.

References

- P. Sun, X. Lu, C. Xu, W. Sun, B. Pan, Understanding of COVID-19 based on current evidence, *J. Med. Virol.* 92 (2020) 548–551, <https://doi.org/10.1002/jmv.25722>.
- G. Cacciapaglia, C. Cot, F. Sannino, Second wave COVID-19 pandemics in Europe: a temporal playbook, *Sci. Rep.* 10 (2020) 15514, <https://doi.org/10.1038/s41598-020-72611-5>.
- G. Fan, Z. Yang, Q. Lin, S. Zhao, L. Yang, D. He, Decreased case fatality rate of COVID-19 in the second wave: a study in 53 countries or regions, *Transbound Emerg Dis* 68 (2021) 213–215, <https://doi.org/10.1111/tbed.13819>.
- R.S. Diaz, T.R.C. Vergara, The COVID-19 second wave: a perspective to be explored, *Braz. J. Infect. Dis.* 25 (2021) 101537, <https://doi.org/10.1016/j.bjid.2020.101537>.
- J. Chen, R. Wang, M. Wang, G.-W. Wei, Mutations strengthened SARS-CoV-2 infectivity, *J. Mol. Biol.* 432 (2020) 5212–5226, <https://doi.org/10.1016/j.jmb.2020.07.009>.
- Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. Liu, L. Nie, H. Qin, M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang, Y. Wang, The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity, *Cell* 182 (2020) 1284–1294, <https://doi.org/10.1016/j.cell.2020.07.012>, e9.
- L. van Dorp, M. Acman, D. Richard, L.P. Shaw, C.E. Ford, L. Ormond, C.J. Owen, J. Pang, C.C.S. Tan, F.A.T. Boshier, A.T. Ortiz, F. Balloux, Emergence of genomic diversity and recurrent mutations in SARS-CoV-2, *Infect. Genet. Evol.* 83 (2020) 104351, <https://doi.org/10.1016/j.meegid.2020.104351>.
- C. Wang, Z. Liu, Z. Chen, X. Huang, M. Xu, T. He, Z. Zhang, The establishment of reference sequence for SARS-CoV-2 and variation analysis, *J. Med. Virol.* 92 (2020) 667–674, <https://doi.org/10.1002/jmv.25762>.
- M. Pachetti, B. Marini, F. Benedetti, F. Giudici, E. Mauro, P. Storic, C. Masciovecchio, S. Angeletti, M. Ciccozzi, R.C. Gallo, D. Zella, R. Ippodrino, Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant, *J. Transl. Med.* 18 (2020) 179, <https://doi.org/10.1186/s12967-020-02344-6>.
- A. Winger, T. Caspari, The spike of concern—the novel variants of SARS-CoV-2, *Viruses* 13 (2021), <https://doi.org/10.3390/v13061002>.
- B. Durmaz, O. Abdulmajed, R. Durmaz, Mutations observed in the SARS-CoV-2 spike glycoprotein and their effects in the interaction of virus with ACE-2 receptor, *Medeni Med J* 35 (2020) 253–260, <https://doi.org/10.5222/MMJ.2020.98048>.
- R. Laskar, S. Ali, Mutational analysis and assessment of its impact on proteins of SARS-CoV-2 genomes from India, *Gene* (2021) 145470, <https://doi.org/10.1016/j.gene.2021.145470>.
- R. Omori, R. Matsuyama, Y. Nakata, The age distribution of mortality from novel coronavirus disease (COVID-19) suggests no large difference of susceptibility by age, *Sci. Rep.* 10 (2020) 16642, <https://doi.org/10.1038/s41598-020-73777-8>.
- O. Albitar, R. Ballouze, J.P. Ooi, S.M. Sheikh Ghadzi, Risk factors for mortality among COVID-19 patients, *Diabetes Res. Clin. Pract.* 166 (2020) 108293, <https://doi.org/10.1016/j.diabres.2020.108293>.
- R. Laskar, S. Ali, Phylo-geo-network and haplogroup analysis of 611 novel coronavirus (SARS-CoV-2) genomes from India, *Life Sci. Alliance* 4 (2021), e202000925, <https://doi.org/10.26508/lsa.202000925>.
- J. Bendl, J. Stourac, O. Salanda, A. Pavelka, E.D. Wieben, J. Zendluka, J. Brezovsky, J. Damborsky, PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations, *PLoS Comput. Biol.* 10 (2014), e1003440, <https://doi.org/10.1371/journal.pcbi.1003440>.
- C.-W. Chen, J. Lin, Y.-W. Chu, iStable: off-the-shelf predictor integration for predicting protein stability changes, *BMC Bioinf.* 14 (2013) S5, <https://doi.org/10.1186/1471-2105-14-S2-S5>.
- S.L. Klein, K.L. Flanagan, Sex differences in immune responses, *Nat. Rev. Immunol.* 16 (2016) 626–638, <https://doi.org/10.1038/nri.2016.90>.
- B. Batista, D. Dickenson, K. Gurski, M. Kebe, N. Rankin, Minimizing disease spread on a quarantined cruise ship: a model of COVID-19 with asymptomatic infections, *Math. Biosci.* 329 (2020) 108442, <https://doi.org/10.1016/j.mbs.2020.108442>.
- M. Gandhi, C. Beyrer, E. Goosby, Masks do more than protect others during COVID-19: reducing the inoculum of SARS-CoV-2 to protect the wearer, *J. Gen. Intern. Med.* 35 (2020) 3063–3066, <https://doi.org/10.1007/s11606-020-06067-8>.
- C. Azgari, Z. Kilinc, B. Turhan, D. Circi, O. Adebali, The mutation profile of SARS-CoV-2 is primarily shaped by the host antiviral defense, *Viruses* 13 (2021), <https://doi.org/10.3390/v13030394>.
- D. Mercatelli, F.M. Giorgi, Geographic and genomic distribution of SARS-CoV-2 mutations, *Front. Microbiol.* 11 (2020) 1800, <https://doi.org/10.3389/fmicb.2020.01800>.
- S. Laha, J. Chakraborty, S. Das, S.K. Manna, S. Biswas, R. Chatterjee, Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission, *Infect. Genet. Evol.* 85 (2020) 104445, <https://doi.org/10.1016/j.meegid.2020.104445>.
- D. Zella, M. Giovanetti, E. Cella, A. Borsetti, M. Ciotti, G. Ceccarelli, G. D'Ettoire, A. Pezzuto, V. Tambone, L. Campanozzi, M. Magheri, F. Unali, M. Bianchi, F. Benedetti, S. Pascarella, S. Angeletti, M. Ciccozzi, The importance of genomic analysis in cracking the coronavirus pandemic, *Expert Rev. Mol. Diagn.* 21 (2021) 547–562, <https://doi.org/10.1080/14737159.2021.1917998>.
- C. Jiang, X. Li, C. Ge, Y. Ding, T. Zhang, S. Cao, L. Meng, S. Lu, Molecular detection of SARS-CoV-2 being challenged by virus variation and asymptomatic infection, *Journal of Pharmaceutical Analysis* (2021), <https://doi.org/10.1016/j.jpfa.2021.03.006>, S2095177921000265.
- M. Bianchi, A. Borsetti, M. Ciccozzi, S. Pascarella, SARS-CoV-2 ORF3a: mutability and function, *Int. J. Biol. Macromol.* 170 (2021) 820–826, <https://doi.org/10.1016/j.ijbiomac.2020.12.142>.
- Z. Cao, H. Xia, R. Rajsbaum, X. Xia, H. Wang, P.-Y. Shi, Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response, *Cell. Mol. Immunol.* 18 (2021) 746–748, <https://doi.org/10.1038/s41423-020-00603-6>.
- M. Rastogi, N. Pandey, A. Shukla, S.K. Singh, SARS coronavirus 2: from genome to infectome, *Respir. Res.* 21 (2020) 318, <https://doi.org/10.1186/s12931-020-01581-z>.
- A. Lopez-Rincon, A. Tonda, L. Mendoza-Maldonado, E. Claassen, J. Garssen, A. D. Kraneveld, A missense mutation in SARS-CoV-2 potentially differentiates between asymptomatic and symptomatic cases, *nCoV* (2020), <https://doi.org/10.2471/BLT.20.258889>.
- F. Caccuri, A. Zani, S. Messali, M. Giovanetti, A. Bugatti, G. Campisi, F. Filippini, E. Scaltriti, M. Ciccozzi, S. Fiorentini, A. Caruso, A persistently replicating SARS-CoV-2 variant derived from an asymptomatic individual, *J. Transl. Med.* 18 (2020) 362, <https://doi.org/10.1186/s12967-020-02535-1>.
- F.K. Yoshimoto, The proteins of Severe Acute respiratory Syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19, *Protein J.* 39 (2020) 198–216, <https://doi.org/10.1007/s10930-020-09901-4>.
- L. Duan, Q. Zheng, H. Zhang, Y. Niu, Y. Lou, H. Wang, The SARS-CoV-2 spike glycoprotein biosynthesis, structure, function, and antigenicity: implications for the design of spike-based vaccine immunogens, *Front. Immunol.* 11 (2020) 576622, <https://doi.org/10.3389/fimmu.2020.576622>.
- S. Lu, Q. Ye, D. Singh, Y. Cao, J.K. Diedrich, J.R. Yates, E. Villa, D.W. Cleveland, K. D. Corbett, The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein, *Nat. Commun.* 12 (2021) 502, <https://doi.org/10.1038/s41467-020-20768-y>.
- T. Shu, M. Huang, D. Wu, Y. Ren, X. Zhang, Y. Han, J. Mu, R. Wang, Y. Qiu, D.-Y. Zhang, X. Zhou, SARS-Coronavirus-2 Nsp13 possesses NTPase and RNA helicase activities that can be inhibited by bismuth salts, *Virol. Sin.* 35 (2020) 321–329, <https://doi.org/10.1007/s12250-020-00242-1>.
- Y. Ma, L. Wu, N. Shaw, Y. Gao, J. Wang, Y. Sun, Z. Lou, L. Yan, R. Zhang, Z. Rao, Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex, *Proc. Natl. Acad. Sci. U.S.A.* 112 (2015) 9436–9441, <https://doi.org/10.1073/pnas.1508686112>.
- S.K. Mishra, T. Tripathi, One year update on the COVID-19 pandemic: where are we now? *Acta Trop.* 214 (2021) <https://doi.org/10.1016/j.actatropica.2020.105778>, 105778.
- R.K. Suryawanshi, R. Koganti, A. Agelidis, C.D. Patil, D. Shukla, Dysregulation of cell signaling by SARS-CoV-2, *Trends Microbiol.* 29 (2021) 224–237, <https://doi.org/10.1016/j.tim.2020.12.007>.
- Y.-T. Tseng, C.-H. Chang, S.-M. Wang, K.-J. Huang, C.-T. Wang, Identifying SARS-CoV membrane protein amino acid residues linked to virus-like particle assembly, *PLoS One* 8 (2013), e64013, <https://doi.org/10.1371/journal.pone.0064013>.
- M. Sarkar, S. Saha, Structural insight into the role of novel SARS-CoV-2 E protein: a potential target for vaccine development and other therapeutic strategies, *PLoS One* 15 (2020), e0237300, <https://doi.org/10.1371/journal.pone.0237300>.
- X. Lei, X. Dong, R. Ma, W. Wang, X. Xiao, Z. Tian, C. Wang, Y. Wang, L. Li, L. Ren, F. Guo, Z. Zhao, Z. Zhou, Z. Xiang, J. Wang, Activation and evasion of type I interferon responses by SARS-CoV-2, *Nat. Commun.* 11 (2020) 3810, <https://doi.org/10.1038/s41467-020-17665-9>.
- R.N. Kirchdoerfer, A.B. Ward, Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors, *Nat. Commun.* 10 (2019) 2342, <https://doi.org/10.1038/s41467-019-10280-3>.
- B. Krichel, S. Falke, R. Hilgenfeld, L. Redecke, C. Uetrecht, Processing of the SARS-CoV pp1a/ab nsp7–10 region, *Biochem. J.* 477 (2020) 1009–1019, <https://doi.org/10.1042/BCJ20200029>.