Journal of Cellular Biotechnology 7 (2021) 85–98
DOI 10.3233/JCB-210037
IOS Press

Point mutation of COVID-19 proteins: A study on noval corona virus (nCov) correlation with MERS and H1N1 viruses and *in silico* investigation of nCoV proteins for future applications

Akansha Kaushik, Swati Kasana, Priyanka Verma and Anuj Kumar Singh*
Department of Biotechnology, Meerut Institute of Engineering and Technology, Meerut, India

Received 4 June 2021
Accepted 1 September 2021

Abstract. Coronavirus disease (COVID 19) which is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2) was first reported in Wuhan, China in December 2019. The disease transformed to a pandemic and affected people's lives all over the world. It caused death to millions of people all over the world. In this project we focused on finding out the correlation of SARS-CoV2 with other respiratory diseases causing viruses like MERS and H1N1 influenza viruses. We further investigated to understand the mutations that occur in the sequences of the SARS-CoV2 during the spread of the disease and correlated it with the functional domains of proteins. The resulted phylogenetic tree indicated that SARS-CoV2 is closely related to the MERS and H1N1 viruses are distantly related. The mutation analysis of 10 different proteins of the SARS-CoV2 shows that there were more than 50 point-mutations among 34 countries sequences for six proteins. Interestingly, four proteins did not any mutation during the analysis. Therefore, these four proteins may be taken into consideration during the development of the diagnostics or therapeutics against this disease.

Keywords: Coronavirus, COVID19, Point mutation, MERS, SARS, Influenza virus, H1N1, nCoV

1. Introduction

The COVID-19 is the biggest pandemic ever heard due to any kind of disaster. The disease was born around the end of December 2019, in the city of Wuhan in China. The name of the disease was due to the virus type, Coronavirus, in the year 2019 (COVID-19), and the causative virus was identified as Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) [1]. Coronaviruses are a group of viruses that cause upper respiratory infection in mammals, birds and develop lethal condition in humans [2]. In the nine month period COVID-19 reached to more than 200 countries and infected more than 36 million and caused 1 million death worldwide [3]. However, the family has historical mark on the globe as another common cold like pandemic with massive death in last two decades.

The symptoms of the novel coronavirus (nCoV) infection has some similarity with SARS, MERS (Middle East respiratory syndrome) and H1N1 (influenza virus) related infections as all were associated with respiratory tract infections [4–6]. The nCoV infections were found to be associated with angiotensin-converting enzyme 2 (ACE2) receptor, mostly present in lungs [7]. However, it was

*Corresponding author: Dr. Anuj Kumar Singh, Assistant Professor, Department of Biotechnology, Meerut Institute of Engineering and Technology, Meerut, India. E-mail: anujsingh77@gmail.com.

ISSN 2352-3689/$35.00

recently reported that infection is not limited to lung nevertheless it reached to the abdominal region [8]. Additionally, in severe cases microcirculatory disorders and systemic endothelial dysfunction were reported recently [9–11]. The present report is based on an *in silico* study of nCoV and associated virus genome sequences analysis to understand the relationship with these respiratory tract infecting viruses. However, the genome sequences of nCoV, SARS-CoV, MERS-CoV was compared with bat-CoV (RaTG13) by Zhou et al where they found similarity 96% with nCoV and 76% with SARS-CoV [6, 12]. Additionally, we have compared the all the protein sequences submitted for nCoV on NCBI from all over the world to find the regions where mutation not happened during the spread which is common in viruses. Later we compared with the sequences submitted from India to understand the domains of the viral protein using bioinformatics tools.

2. Methods

2.1. Sequence collection

We collected 34 complete genome sequences of nCoV freshly submitted in NCBI database till 18 March 2020. Additionally, we have collected 4 MERS-CoV sequences of China and U.S from the Viral Genome Database and 9 H1N1 influenza virus sequences (submitted from India and China) from Influenza Virus Database using Open Flu Database (Table 1a & b). Next, we collected all types of proteins sequences for the same nCoV from NCBI which was used for comparison with MERS and HIN1 genome.

2.2. Multiple sequence alignment using clustal omega

The collected genome sequences were used for multiple sequence alignment (MSA) using online tool, Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) which helped us to collect Jalview format of the aligned sequences [13, 14].

2.3. Jalview based analysis of genome and protein sequences

Jalview (https://www.jalview.org/) is a free bioinformatics tool for the analysis of DNA, RNA, and proteins. After performing MSA in clustal omega, we download Jalview format of MSA result and then we use Jalview offline software to visualize the result and export the data in FASTA format [15].

2.4. MEGA (Molecular evolutionary genetics analysis) for phylogenetic analysis

The exported FASTA file of the MSA was opened using an offline tool, MEGA-X [https://www.megasoftware.net/] for phylogenetic analysis and construction of a phylogenetic tree was generated using maximum likelihood method [16].

2.5. InterPro scan database for domain analysis

The InterPro Scan database has the information to understand the protein families and its functional domain. To understand the functional domain of each protein we used an Indian submitted sequence for nCoV (MT012098). We obtained the domain information from the database of individual proteins of the virus and correlated the mutation result [17].

Table 1a
Gene codes and countries for the sources of genome sequences for nCoV

SARS-CoV-2 complete genome	
GenBank ID	Locality
MT007544	Australia: Victoria
MT126808	Brazil
MT135041	China: Beijing_1
MT121215	China: Shanghai_2
MN996527	China: Wuhan_3
MT256924	Colombia: Antioquia
MT020781	Finland
MT012098	India: Kerala State_1
MT050493	India: Kerala State_2
MT281530	Iran
MT276597	Israel_1
MT276598	Israel_2
MT066156	Italy
LC528232	Japan_1
LC528233	Japan_2
LC529905	Japan_3
MT072688	Nepal
MT240479	Pakistan: Gilgit_1
MT262993	Pakistan: KPK_2
MT263074	Peru
MT039890	South Korea
MT198652	Spain: Valencia_1
MT233519	Spain: Valencia_2
MT233520	Spain: Valencia_3
MT093571	Sweden
MT066175	Taiwan_1
MT066176	Taiwan_2
MT192759	Taiwan_3
MN994467	USA: CA_1
MT276329	USA: FL_2
MT106054	USA: TX_3
MN985325	USA: WA_4
MT192772	Viet Nam: Ho Chi Minh city_1
MT192773	Viet Nam: Ho Chi Minh city_2

3. Results

3.1. Phylogenetic analysis of nCoV with MERS-CoV and H1N1

To understand the relationship among the nCoV and other respiratory diseases we collected 34 complete genome sequences of coronavirus of various countries which were submitted till 18 March 2020 in the NCBI database. The evolutionary relationships for these nCoV sequences were analysed with 4 MERS sequences and 9 H1N1 influenza virus sequences. The obtained phylogenetic tree revealed that the nCoV is distantly related to H1N1 (influenza virus) and MERS is closely related

Table 1b
Gene codes and countries for the sources of genome sequences for MERS and H1N1 (influenza virus)

MERS-CoV	
GenBank ID	Locality
KT006149	China
KJ813439	USA
KP223131	USA
KJ829365	USA
Influenza viruses (H1N1)	
OFL181342	China: Beijing
OFL180257	China: Beijing
OFL180259	China: Beijing
OFL287088	India: Bangalore
OFL287089	India: Bangalore
OFL287090	India: Bangalore
OFL287092	India: Bangalore
OFL287093	India: Bangalore
OFL287094	India: Bangalore

*OFL - OpenFlu database by Swiss Institute of Bioinformatics.

(Fig. 1). Our data supports the recent study shown its relation with various SARS viruses including MERS [6, 16].

3.2. Mutation analysis using jalview in nCoV proteins

Viruses are known for changing their coat proteins during their life-cycle since it utilises host expression system. Considering the possibility of changes in nCoV associated proteins during the pandemic we compared all the 34 entries for the mutation occurred in the viral proteins. The individual proteins were studied using MSA in CLUSTAL omega followed by Jalview analysis for mutation search. We compared 10 different proteins present in nCoV: orf1ab polyprotein, surface glycoprotein (spike protein), orf3a protein, envelop protein, membrane glycoprotein, orf6 protein, orf7a protein, orf8 protein, nucleocapsid phospho-protein and orf10 protein. Interestingly, we found mutation in the 6 proteins among various countries' submitted sequences. However, no mutations were observed among the 34 countries' sequences for the four viral proteins (membrane glycoprotein, orf6 protein, orf7a protein and orf10 protein) during our analysis (Table 2).

The highest point mutations were observed in orf1ab (35 different positions), orf 8 protein (12 positions) and nucleocapsid phosphor-protein (11 positions) among the sequences used for analysis [20]. Next, we needed to know the protein domains affected by the mutation which directed us to do domain analysis.

3.3. Domain analysis of individual proteins of an Indian sequence of SARS-CoV-2

To understand the domains of individual proteins in nCoV we used the coronavirus sequence submitted from India (with the Acc. no.- MT012098). In order to perform the domain analysis of individual protein of the virus, first we collected the amino acid sequence of individual proteins then sequence of the individual protein was uploaded on the InterPro Scan online tool separately and the results were

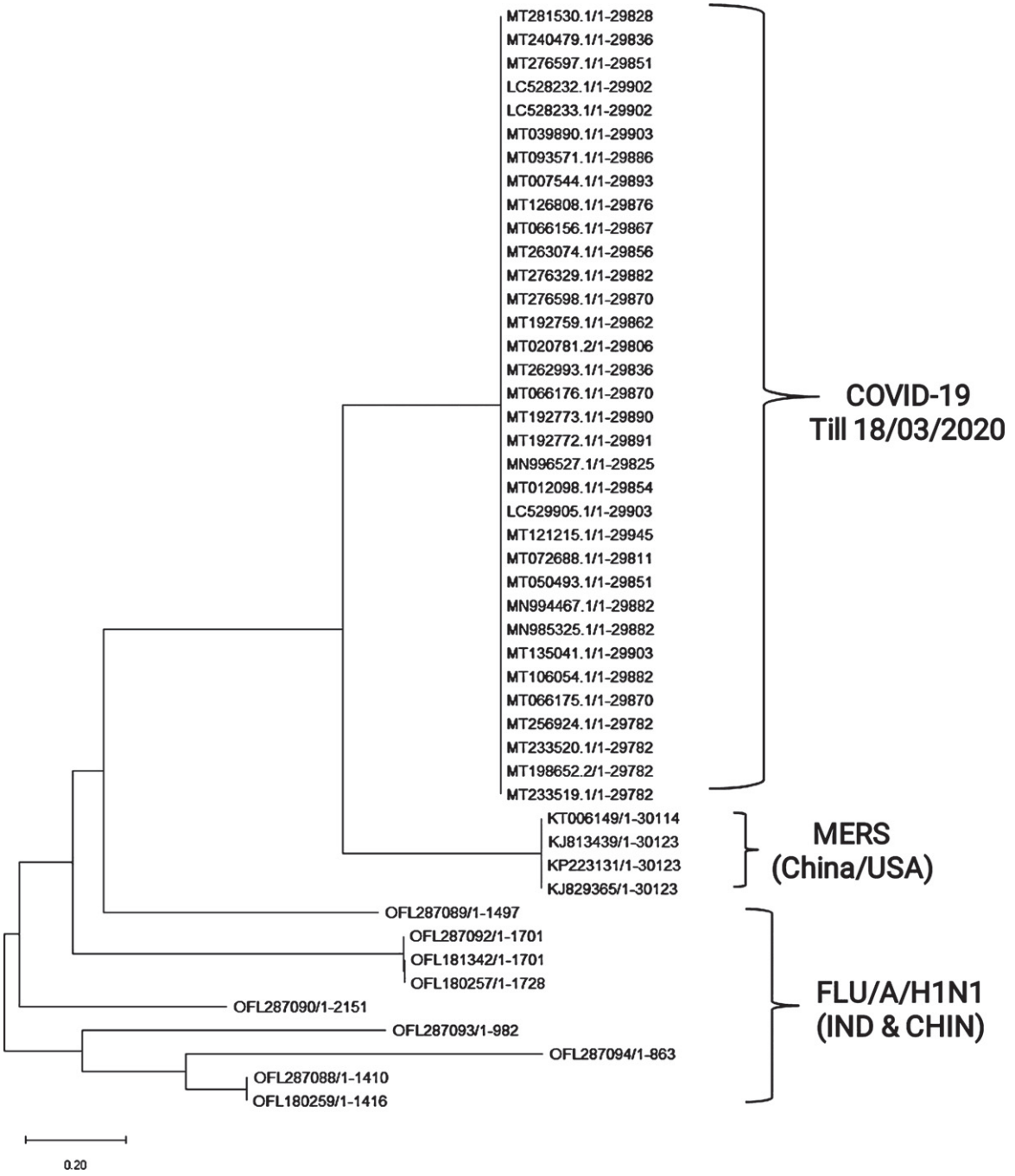

Fig. 1. Phylogenetic analysis of nCoV, MERS and H1N1 (influenza virus): The 34 complete genome sequences of nCoV were compared with 4 MERS sequences and 9 H1N1 sequences in order to build a phylogenetic tree in the MEGA-X software using the Maximum Likelihood method and Tamura-Nei model [18, 19]. In the analysis the codon positions included were 1st, 2nd, 3rd and the non-coding regions. There was a total of 29945 positions in the final dataset. The default settings were used for the analysis.

Table 2
Mutations in different sequences of different countries

Protein	GenBank	Country	Mutation	Location
ofr 1ab	MT240479	PAKISTAN1/1–7096	Arginine to cysteine	207
	MT281530	IRAN/1–7096	Valine to isoleucine	378
	MT240479	PAKISTAN1/1–7096	Valine to isoleucine	378
	MN994467	USA1/1–7096	Serine to asparagine	428
	MT050493	INDIA 2/1–7096	Isoleucine to valine	476
	MT012098	INDIA1/1–7096	Isoleucine to theronine	671
	MT093571	SWEDEN/1–7096	Glycine to serine	818
	MT039890	SOUTH/1–7096	Methionine to isoleucine	902
	MT135041	CHINA1/1–7096	Leucine to phenylalanine	1599
	MT121215	CHINA2/1–7096	Proline to serine	1921
	MT050493	INDIA2/1–7096	Proline to leucine	2079
	MT012098	INDIA1/1–7096	Proline to serine	2144
	MT263074	PERU/1–7096	Asparagine to asparatic acid	2894
	MT240479	PAKISTAN1/1–7096	Proline to leucine	2985
	MT233520	SPAIN3/1–7096	Phenylalanine to tyrosine	3071
	MT198652	SPAIN1/1–7096	Phenylalanine to tyrosine	3071
	MT233519	SPAIN2/1–7096	Phenylalanine to tyrosine	3071
	MT192772	VIETNAM1/1–7096	Arginine to cysteine	3323
	MT192773	VIETNAM2/1–7096	Arginine to cysteine	3323
	MT126808	BRAZIL/1–7096	Leucine to phenylalanine	3606
	MT276597	ISRAEL 1/1–7096	Leucine to phenylalanine	3606
	LC528232	JAPAN 1/1–7096	Leucine to phenylalanine	3606
	LC528233	JAPAN 2/1–7096	Leucine to phenylalanine	3606
	MT240479	PAKISTAN1/1–7096	Leucine to phenylalanine	3606
	MT281530	IRAN1/1–7096	Leucine to phenylalanine	3606
	MT093571	SWEDEN/1–7096	Phenylalanine to leucine	4321
	MT263074	PERU/1–7096	Proline to leucine	4715
	MT276597	ISRAEL1/1–7096	Proline to leucine	4715
	MT276329	USA2/1–7096	Proline to leucine	4715
	MT012098	INDIA1/1–7096	Alanine to valine	4798
	MT050493	INDIA2/1–7096	Threonine to isoleucine	5540
	MT281530	IRAN/1–7098	Threonine to isoleucine	6040
	MT106054	USA3/1–7096	Aspartic acid to alanine	6306
	MT039890	SOUTH/1–7096	Threonine to methionine	6893
	MN996527	CHINA3/1–7096	Aspartic acid to asparagine	7020
Surface glycoprotein	MT039890	SOUTH/1–75	Leucine to histidine	37
orf3a protein	MT281530	IRAN/1–275	Tryptophan to leucine	128
	LC529905	JAPAN3/1–275	Leucine to valine	140
	MT198652	SPAIN1/1–275	Glycine to valine	196
	MT233519	SPAIN2/1–275	Glycine to valine	196
	MT233520	SPAIN3/1–275	Glycine to valine	196
	MT039890	SOUTH/1–275	Glycine to valine	251

(*Continued*)

Table 2
(*Continued*)

Protein	GenBank	Country	Mutation	Location
	MT007544	AUSTRALIA/1–275	Glycine to valine	251
	MT066156	ITALY/1–275	Glycine to valine	251
	MT093571	SWEDEN/1–275	Glycine to valine	251
	MT126808	BRAZIL/1–275	Glycine to valine	251
Envelop protein	MT039890	SOUTH/1–75	Leucine to histidine	37
Membrane glycoprotein		All	NO MUTATION	
orf 6 protein		All	NO MUTATION	
orf7a protein		All	NO MUTATION	
orf 8 protein	MT106054	USA 3/1–121	Threonine to isoleucine	11
	MN994467	USA 1/1–121	Valine to leucine	62
	MN994467	USA 1/1–121	Leucine to serine	84
	MT106054	USA 3/1–121	Leucine to serine	84
	MT135041	CHINA1/1–121	Leucine to serine	84
	MT256924	COLOMBIA/1–121	Leucine to serine	84
	MT050493	INDIA 2/1–121	Leucine to serine	84
	MT198652	SPAIN 1/1–121	Leucine to serine	84
	MT233519	SPAIN 2/1–121	Leucine to serine	84
	MT233520	SPAIN 3/1–121	Leucine to serine	84
	MT066175	TIAWAN 1/1–121	Leucine to serine	84
	MN985325	USA 4/1–121	Leucine to serine	84
Nucleocapsid phosphor-protein	MT198652	SPAIN /1–419	Serine to leucine	197
	MT198652	SPAIN 1/1–419	Serine to leucine	197
	MT233519	SPAIN 2/1–419	Serine to leucine	197
	MT276598	ISRAEL 2/1–419	Arginine to lysine	203
	MT263074	PERU/1–419	Arginine to lysine	203
	MT276329	USA2/1–419	Arginine to lysine	203
	MT276598	ISRAEL 2/1–419	Glysine to arginine	204
	MT263074	PERU/1–419	Glysine to arginine	204
	MT276329	USA2/1–419	Glysine to arginine	204
	MT256924	COLOMIA/1–419	Glysine to cysteine	238
	LC529905	JAPAN3/1–419	Proline to serine	344
orf 10 protein		All	NO MUTATION	

obtained [17, 21]. The obtained results for all 10 proteins were represented in Table 3 along with the gene ontology (GO).

The proteins of SARS-CoV-2 of an Indian sequence in which the domains are present are as follows:

orf1ab polyprotein was the largest protein with 20 domains, surface glycoprotein has two domains but other proteins (orf3a, M-protein, orf7a, orf8 and nucleocapsid phosphor-protein) has one domain each. Interestingly, we did not observe any domains in the analysis of envelope (E) protein, orf6 protein and orf10 protein.

The domain analysis of one submission (MT012098) for SARS-CoV-2 revealed the information about the domains of nCoV proteins. Later the MSA based mutation analysis results were mapped with domain analysis results considering all 34 entries must have similar domain distributions. The mapped results were represented in Table 4.

Table 3
Different proteins and domains in the SARS-CoV2 sequences

Gene code	Protein name	Domain name and IPR code	Amino acid range	Functions & Gene Ontology (GO)
QHS34545.1	ORF 1ab	NSP 1 (IPR02590)	13–127	Viral genome replication (GO:0019079)
		SARS-CoV_Nsp3_N (IPR0024358)	920–986	Transcription, DNA-templated (GO:0006351)
		Macro_dom (IPR002589)	1025–1194	Viral protein processing (GO:0019082)
		Nsp3_PLR2pro (IPR022733)	1498–1561	Viral RNA genome replication (GO:0039694)
		Nsp3_coronavir (IPR024375)	1351–1493	Proteolysis (GO:0006508)
		Viral_protease (IPR014827)	1564–1882	Transferase activity (GO:0016740)
		Peptidase_C30/C16 (IPR013016)	1634–1898	
		NAR_dom (IPR032592)	1922–2019	Cysteine–type peptidase activity (GO:0008234)
		Corona_NSP4_C (IPR032505)	3166–3261	Nucleic acid binding (GO:0003676)
		Peptidase_C30 (IPR008740)	3264–3582	Zinc ion binding (GO:0008270)
		NPS7 (IPR014828)	3860–3942	RNA-directed 5'-3' RNA polymerase activity (GO:0003968)
		NSP8 (IPR014829)	3943–4140	ATP binding (GO:0005524)
		NSP9 (IPR014822)	4141–4253	Cysteine-type endopeptidase activity (GO:0004197)
		RNA_synth_NSP10_coronavirus (IPR018995)	4262–4384	
		RNA_pol_N_coronovir (IPPR009469)	4407–4758	RNA binding (GO:0003723)
		RNA-dir_pol_Psvirus (IPR007094)	5004–5166	

(*Continued*)

Table 3
(*Continued*)

Gene code	Protein name	Domain name and IPR code	Amino acid range	Functions & Gene Ontology (GO)
		CV_ZBD (IPR027352)	5325–5408	
		(+)RNA_virus_helicase_core_dom (IPR027351)	5581–5932	Methyltransferase activity (GO:0008168)
		NSP11 (IPR009466)	5929–6520	Exoribonuclease activity, producing 5'-phosphomonoesters (GO:0016896)
		Coronavirus_NSP16 (IPR009461)	6800–7095	Omega peptidase activity (GO:0008242)
QHS34546.1	S-protein	Spike_rcpt_bd (IPR018548)	285–538	Membrane fusion (GO:0061025)
		Corona_S2 (IPR002552)	641–1225	Receptor-mediated virion attachment to host cell (GO:0046813
QHS34547.1	ORF 3a	SARS_Coronavirus_Orf3/3a (IPR024407)	1–230	
QHS34548.1	E-protein	NO DOMAIN		
QHS34549.1	M-protein	Corona_M (IPR002574)	1–177	Viral life cycle (GO:0019058)
QHS34550.1	ORF 6	NO DOMAIN		
QHS34551.1	ORF 7	SARS_X4 (IPR014888)	1–054	
QHS34552.1	ORF 8	Corona_NS8 (IPR022722)	1–074	
QHS34553.1	ORF 9	Corona_nucleocap (IPR001218)	1–374	Viral nucleocapsid (GO:0019013)

4. Discussion and conclusion

In our study, we compared the genome sequences of upper respiratory tract infecting viruses to check the relationship with nCoV. In the present study we also compared various proteins of nCoV to find out the mutation during spread of the disease. The overall finding suggest that the nCoV belong to the same family which caused SARS and MERS like pandemic earlier in small part of the world [2]. The mutation analysis suggested that the highest number (10) of mutation was found in orf8 protein where leucine was mutated to serine in counties like -USA, India, Spain and China but all these are at the region which does not belong to any functional domain of the protein. Next was glycine to valin in orf3

Table 4
Mappinng of mutation analysis data with the domain analysis data

Protein	GenBank	Mutation	Location	Predicted domain
ORF 1ab	MT240479	Arginine to cysteine	207	No domain predicted
	MT281530	Valine to isoleucine	378	No domain predicted
	MT240479	Valine to isoleucine	378	No domain predicted
	MN994467	Serine to asparagine	428	No domain predicted
	MT050493	Isoleucine to valine	476	No domain predicted
	MT012098	Isoleucine to theronine	671	No domain predicted
	MT093571	Glycine to serine	818	No domain predicted
	MT039890	Methionine to isoleucine	902	No domain predicted
	MT135041	Leucine to phenylalanine	1599	Viral protease
	MT121215	Proline to serine	1921	No domain predicted
	MT050493	Proline to leucine	2079	No domain predicted
	MT012098	Proline to serine	2144	No domain predicted
	MT263074	Asparagine to asparatic acid	2894	No domain predicted
	MT240479	Proline to leucine	2985	No domain predicted
	MT233520	Phenylalanine to tyrosine	3071	No domain predicted
	MT198652	Phenylalanine to tyrosine	3071	No domain predicted
	MT233519	Phenylalanine to tyrosine	3071	No domain predicted
	MT192772	Arginine to cysteine	3323	Peptidase_C30/C16
	MT192773	Arginine to cysteine	3323	Peptidase_C30/C16
	MT126808	Leucine to phenylalanine	3606	No domain predicted
	MT276597	Leucine to phenylalanine	3606	No domain predicted
	LC528232	Leucine to phenylalanine	3606	No domain predicted
	LC528233	Leucine to phenylalanine	3606	No domain predicted
	MT240479	Leucine to phenylalanine	3606	No domain predicted
	MT281530	Leucine to phenylalanine	3606	No domain predicted
	MT093571	Phenylalanine to leucine	4321	RNA-syn-NSP10-coronavirus
	MT263074	Proline to leucine	4715	RNA_pol_N_coronovir
	MT276597	Proline to leucine	4715	RNA_pol_N_coronovir
	MT276329	Proline to leucine	4715	RNA_pol_N_coronovir
	MT012098	Alanine to valine	4798	No domain predicted
	MT050493	Threonine to isoleucine	5540	No domain predicted
	MT281530	Threonine to isoleucine	6040	NSP11
	MT106054	Aspartic acid to alanine	6306	No domain predicted
	MT039890	Threonine to methionine	6893	NSP16
	MN996527	Aspartic acid to asparagine	7020	NSP16
Surface glycoprotein	MT039890	Leucine to histidine	37	No domain predicted
Orf 3a	MT281530	Tryptophan to leucine	128	No domain predicted
	LC529905	Leucine to valine	140	No domain predicted

(*Continued*)

Table 4
(*Continued*)

Protein	GenBank	Mutation	Location	Predicted domain
	MT198652	Glycine to valine	196	No domain predicted
	MT233519	Glycine to valine	196	No domain predicted
	MT233520	Glycine to valine	196	No domain predicted
	MT039890	Glycine to valine	251	No domain predicted
	MT007544	Glycine to valine	251	No domain predicted
	MT066156	Glycine to valine	251	No domain predicted
	MT093571	Glycine to valine	251	No domain predicted
	MT126808	Glycine to valine	251	No domain predicted
Envelop protein	MT039890	Leucine to histidine	37	No domain predicted
Membrane glycoprotein		NO MUTATION		
orf 6 protein		NO MUTATION		
orf7a protein		NO MUTATION		
orf 8	MT106054	Threonine to isoleucine	11	Corona_NS8
	MN994467	Valine to leucine	62	Corona_NS8
	MN994467	Leucine to serine	84	No domain predicted
	MT106054	Leucine to serine	84	No domain predicted
	MT135041	Leucine to serine	84	No domain predicted
	MT256924	Leucine to serine	84	No domain predicted
	MT050493	Leucine to serine	84	No domain predicted
	MT198652	Leucine to serine	84	No domain predicted
	MT233519	Leucine to serine	84	No domain predicted
	MT233520	Leucine to serine	84	No domain predicted
	MT066175	Leucine to serine	84	No domain predicted
	MN985325	Leucine to serine	84	No domain predicted
Nucleocapsid phosphor-protein	MT198652	Serine to leucine	197	Corona_nucleoca
	MT198652	Serine to leucine	197	Corona_nucleoca
	MT233519	Serine to leucine	197	Corona_nucleoca
	MT276598	Arginine to lysine	203	Corona_nucleoca
	MT263074	Arginine to lysine	203	Corona_nucleoca
	MT276329	Arginine to lysine	203	Corona_nucleoca
	MT276598	Glysine to arginine	204	Corona_nucleoca
	MT263074	Glysine to arginine	204	Corona_nucleoca
	MT276329	Glysine to arginine	204	Corona_nucleoca
	MT256924	Glysine to cysteine	238	No domain predicted
	LC529905	Proline to serine	344	No domain predicted
orf 10 protein		NO MUTATION		

protein (8) among Spain, South Korea, Australia, Italy, Sweden and Brazil submitted nCoV sequences at unpredictable domains. The similar analysis we did for various point mutations in the given table below (Table 5). Finding the significance of these mutations can be correlated with the severity of cases in certain countries. However, for identification of new targetable proteins those proteins can be used which did not show any mutation.

Table 5
Comparative analysis of point mutations

Type of mutation	No. of mutation	Protein	Country
Leucine to serine	10	orf 8	USA China Colombia India Spain Taiwan
Glycine to valine	8	orf 3a	Spain South Korea Australia Italy Sweden Brazil
Leucine to phenylalanine	7	Orf 1ab	China Brazil Israel Japan Pakistan Iran
Proline to leucine	5	Orf 1ab	India Pakistan Peru USA
Proline to serine	3	Orf 1ab Nucleocapsid phosphor-protein	China India Japan
Phenyalanine to tyrosine	3	Orf 1ab	Spain
Threonine to isoleucine	3	Orf 1ab Orf 8	India Iran USA
Serine to leucine	3	Nucleocapsid phosphor-protein	Spain
Arginine to lysine	3	Nucleocapsid phosphor-protein	Israel Peru USA
Glycine to arginine	3	Nucleocapsid phosphor-protein	Israel Peru USA
Arginine to cysteine	3	Orf 1ab	Pakistan Veitnam
Valine to isoleucine	2	Orf 1ab	Iran Pakistan
Leucine to histidine	2	Surface glycoprotein Envelop protein	South Korea South Korea

(*Continued*)

Table 5
(*Continued*)

Type of mutation	No. of mutation	Protein	Country
Serine to asparagine	1	Orf 1ab	USA
Isoleucine to valine	1	Orf 1ab	India
Isoleucine to threonine	1	Orf 1ab	India
Glycine to serine	1	Orf 1ab	Sweden
Methionine to isoleucine	1	Orf 1ab	South
Asparagine to aspartic acid	1	Orf 1ab	Peru
Alanine to valine	1	Orf 1ab	India
Aspartic acid to alanine	1	Orf 1ab	USA
Threonine to methionine	1	Orf 1ab	South
Aspartic acid to asparagine	1	Orf 1ab	China
Tryptophan to leucine	1	Orf 3a	Iran
Leucine to valine	1	Orf 3a	Japan
Valine to leucine	1	Orf 8	USA
Glycine to cysteine	1	Nucleocapsid phosphor-protein	Colombia

Conflicts of interest

The authors have no conflict of interest to report.

References

[1] COVID-19. European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/covid-19-pandemic. Accessed 30 Apr 2020.
[2] Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. Virol J. 2019;16:69.
[3] Coronavirus Update (Live): 36,792,906 Cases and 1,067,469 Deaths from COVID-19 Virus Pandemic Worldometer. [https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?%22. Accessed 9 Oct 2020.]
[4] Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. Lancet Infect Dis. 2020;20:e238-44.
[5] He D, Zhao S, Li Y, Cao P, Gao D, Lou Y, et al. Comparing COVID-19 and the 1918–19 influenza pandemics in the United Kingdom. Int J Infect Dis. 2020;98:67-70.
[6] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579:270-3.
[7] Ni W, Yang X, Yang D, Bao J, Li R, Xiao Y, et al. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. Crit Care. 2020;24:422.
[8] Jung EM, Stroszczynski C, Jung F. Contrast enhanced ultrasound (CEUS) to assess pleural pulmonal changes in severe COVID-19 infection: First results. Clin Hemorheol Microcirc. 2020;75:19-26.
[9] Jung F, Krüger-Genge A, Franke RP, Hufert F, Küpper J-H. COVID-19 and the endothelium. Clin Hemorheol Microcirc. 2020;75:7-11.
[10] Martini R. The compelling arguments for the need of microvascular investigation in COVID-19 critical patients. Clin Hemorheol Microcirc. 2020;75:27-34.
[11] Jung EM, Stroszczynski C, Jung F. Contrast enhanced ultrasonography (CEUS) to detect abdominal microcirculatory disorders in severe cases of COVID-19 infection: First experience. Clin Hemorheol Microcirc. 2020;74:353-61.
[12] Singh AK, Chaterjee A, Sirohi S, Sharma N, Kathuria A. Convalescent plasma therapy: A promising solution for SARS-CoV-2 outbreak. J Cell Biotechnol. 2021;7:11-7.
[13] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.

[14] Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018;27:135-45.
[15] Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinforma Oxf Engl. 2009;25:1189-91.
[16] Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. J Med Virol. 2020;92:522-8.
[17] Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 2019;47:D351-60.
[18] Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10:512-26.
[19] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35:1547-9.
[20] Yu Y, Santat LA, Choi S. 6 - Bioinformatics Packages for Sequence Analysis. In: Arora DK, Berka RM, Singh GB, editors. Applied Mycology and Biotechnology. Elsevier; 2006; p. 143-60.
[21] Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 2017;45:D190-9.