Erzincan Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
2020, 13(3), 1187-1197
ISSN: 1307-9085, e-ISSN: 2149-4584
Araştırma Makalesi

Erzincan University
Journal of Science and Technology
2020, 13(3), 1187-1197
DOI: 10.18185/erzifbed.758406
Article Research

Comparative Genomic and Proteomic Analysis of SARS CoV-2 - with Potential Mutation Probabilities and Drug Targeting

Ekrem AKBULUT[1*]

[1]Department of Bioengineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Özal University, Malatya, Turkey

Geliş / Received: 26/06/2020, Kabul / Accepted: 22/12/2020

Abstract

COVID-19 caused by the highly pathogenic SARS-CoV-2 has caused the death of over 1.69 million people worldwide. High mutation potentials of RNA viruses require the determination of the most accurate structure to be targeted for treatment. In this study, comparative genomic and proteomic analyses of SARS-CoV-2 were performed using SARS-CoV and MERS-CoV, and the mutation potential of the residues was analyzed using bioinformatics tools. SARS-CoV-2 was found to be 80.08% and 58.79% similar to SARS-CoV and MERS-CoV, respectively, at the nucleotide level. G+C content were 38%, 40.8% and 41.2% for SARS-CoV-2, SARS-CoV and MERS-CoV, respectively. 5′UTR G+C content was 44.6%, 43.5% and 44.7% for SARS-CoV-2, MERS-CoV and SARS-CoV, respectively. At the amino acid level, SARS-CoV-2 and SARS-CoV showed 83.3% similarity, whereas SARS-CoV-2 and MERS-CoV showed 42.5% similarity. The E, M, N and S proteins of SARS-CoV-2 and SARS-CoV were found to be 94%, 90.1%, 90.6% and 76.1% identical, respectively. For SARS-CoV-2, 14 residues with a high risk of mutation and their repeat numbers in the genome were identified. Data from this study reveal that non-functional conserved proteins such as ORF6 and ORF7b with low risk of mutation may be appropriate targets for the treatment because of their functional properties.

Keywords: SARS CoV-2, COVID-19, genome, proteome, mutation

SARS CoV-2'nin Karşılaştırmalı Genomik ve Proteomik Analizi – İlaç Hedefleme ve Potansiyel Mutasyon Olasılıkları

Öz

Yüksek oranda patojenik SARS-CoV-2'nin neden olduğu COVID-19 dünya genelinde 1.69 milyondan fazla insanın ölümüne neden oldu. RNA virüslerinin yüksek mutasyon potansiyelleri tedavi için en doğru yapının tanımlanmasını gerektirir. Bu çalışmada COVID19'un aynı alt sınıfta yer alan SARS ve MERS ile karşılaştırmalı genomik, proteomik analizleri ve rezidülerin mutasyon potansiyelleri biyoinformatik araçlar ile analiz edildi. COVID19'un nükleotid düzeyinde SARS ile 80.08% ve MERS ile 58.79% benzer olduğu bulundu. GC% oranları COVID19, SARS ve MERS sırası ile 38%, 40.8% ve 41.2%'dir. 5'UTR GC içeriği COVID19 (44.6%), MERS (43.5%) ve SARS (44.7%)'dir. Aminoasit düzeyinde COVID19, SARS ile 83.3%, MERS ile 42.5% benzerlik gösterdi. Temel yapısal proteinler kıyaslandığında COVID19/SARS'ın E, M, N ve S-proteinleri sırasıyla 94.7, 90.1, 90.6 and 76,1% oranında aynıdır. COVID19 için 14 yüksek mutasyon riski olan rezidü ve genomda tekrar sayıları belirlendi. Sonuç olarak; COVID19 ve SARS'ın yüksek yapısal benzerlikleri proteinlerin fonksiyonel benzerliklerine işaret edebilir. Bu çalışmanın verileri, düşük mutasyon riski ile ORF6 ve ORF7b gibi fonksiyonel olmayan korunmuş proteinlerin fonksiyonel özellikleri ile tedavi için uygun hedefler olabileceğini işaret etmektedir.

Anahtar Kelimeler: SARS CoV-2, COVID-19, genom, proteom, mutasyon

1. Introduction

Betacoronaviruses (β-CoV) are enveloped, positive-sense, single-stranded RNA viruses of zoonotic origin, belonging to the Coronaviridae family of the order Nidovirales (Baltimore, 1971; Weiss and Leibowitz, 2011; Huang et al, 2020a). Coronaviruses, which were described in humans in the 1960s and characterised by cold symptoms, have come to the fore with their fatal effect over the last 20 years (Al-Osail and Al-Wazzah, 2017). The severe acute respiratory syndrome (SARS) epidemic, which began in 2002 in the Guangdong Province of China and spread to five continents with a total of 8,098 infected cases and 774 deaths, was followed by the Middle-East respiratory syndrome (MERS) epidemic that appeared in Saudi Arabia in 2012, which resulted in a total of 2494 infected cases from 27 countries and 858 deaths (Drosten et al., 2003; Zaki et al., 2012). The most prominent strain of the betacoronavirus family on a global scale is that of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which emerged in the city of Wuhan in Hubei Province of China in 2019, resulting in a pandemic with more than 76.8 million infected cases and over 1.69 million deaths (Chan et al., 2020; Worldometer, 2020; Zhu et al., 2020).

Apart from the highly pathogenic coronaviruses such as SARS-CoV-2, MERS-CoV and SARS-CoV, four endemic strains with low pathogenicity (HCoV-OC43, HCoV-229E, HCoV-NL63 and HCoV-KU1) are known (McBride and Fielding, 2012; Walls et al, 2020). These differ from each other in their genetic structure and antigenic properties. Although there are different hypotheses about intermediate hosts, it is considered to have originated in bats due to high genetic similarities (Lam et al., 2020; Li et al., 2020). It is important to quickly reveal the structural characteristics of SARS-CoV-2 because of its virulence and the number of deaths it has caused worldwide.

Understanding the genetic and proteomic structure is important for studies on drug and vaccine design in the fight against diseases (Abubucker et al., 2011; Pardi et al., 2018; Shereen et al., 2020). Similarities/differences, topological and physicochemical properties in the genome and proteome structure contribute to defining the target structure for treatment (Badani et al., 2014; Liu et al., 2020; Mahajan et al., 2018). The stability of the genetic and protein structure to be targeted in the studies determines the validity of the developed treatment. There is also a risk that a mutation in the genome will invalidate the developed treatment (Regla-Nava et al., 2015; Shen et al., 2003). In this study, SARS-CoV-2 was compared with two other virulent members of the betacoronavirus family (SARS-CoV and MERS-CoV) at the genome and proteome level. Similarities, physicochemical and topological features, and mutational risks at the residue level of SARS-CoV-2 were analyzed using bioinformatics tools to understand the target structure for treatment.

2. Material and Methods

Sequence information of SARS (NC_004718.3), MERS (NC_019843.3) and COVID19 (NC_045512.2) were accessed from the NCBI database. Sequences were aligned with the FFT-NS-i strategy using the

MAFFT (Version 7.452) multiple sequence alignment program (Carroll et al., 2007; K. Katoh, 2002; Kazutaka Katoh et al., 2018). sequences were conducting using the Poisson correction model (Zuckerkandl & Pauling, 1965). The homogeneity of substitution patterns between sequences and estimates of net base composition bias disparity between sequences were conducted with Disparity index test. A Monte Carlo test (500 replicates) was used to estimate the P-values (Kumar and Gadagkar, 2001) which smaller than 0.05 are considered significant. Amino acid exchange probability and mutation data matrices was analyzed by JTT mutation model (Jones et al., 1992). Physicochemical properties of protein sequences were analyzed Emboss PepInfo (McWilliam et al., 2013). GC% contents pf nucleotide sequences were calculate Mega X (Kumar et al., 2018). Amino acid compositions of protein sequences were analyzed with PepStats (Rice et al., 2000). The hydropathy index was calculated with Expasy ProtScale for possible structural properties (Wilkins et al., 1999). Bioinformatics analysis was supported with MegaX (Kumar et al., 2018).

3. Results and Discussion

All preventive or therapeutic drugs against diseases target either the DNA or protein Estimates of evolutionary divergence between

structure (Overington et al., 2006). The SARS-CoV-2 genome consists of 12 protein encoding regions of size 29.9 kb. SARS-CoV and MERS-CoV genomes consist of 14 and 11 encoding regions of size 29.7 kb and 30.1 kb, respectively.

In this study, SARS-CoV-2 was found to be 80.08% and 58.79% similar to SARS-CoV and MERS-CoV, respectively, at the nucleotide level using the MAFFT FFT-NS-i strategy. The similarity ratio between SARS-CoV-2 and MERS-CoV increased within the range of 13–21 kb of the genome (Figure 1). High identity at the nucleotide level between SARS-CoV-2 and SARS-CoV indicates similar structural and physicochemical properties. Although the SARS-CoV-2 genome has not been fully elucidated, computational studies have been recently published that show high similarity between the spike and nucleocapsid structures of SARS-CoV and SARS-CoV-2 and support the data presented in this study (Chatterjee, 2020; Ul-Qamar et al, 2020; Walls et al, 2020). This homology between the two strains provides important data for understanding the functional properties of SARS-CoV-2 proteins.

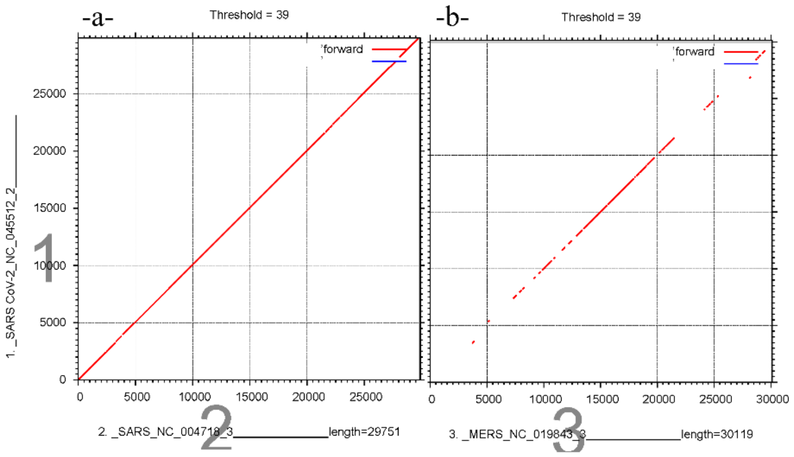

Figure 1. Identity of SARS CoV2, SARS CoV and MERS CoV sequences. a)Comparison SARS CoV-2 with SARS CoV, b) Comparison SARS CoV-2 with MERS CoV

The G+C content for SARS-CoV-2, SARS-CoV and MERS-CoV were 38%, 40.8/ and 41.2%, respectively (Figure 2). The GC content is one an important parameter affecting the stability of the three-dimensional (3D) structure of proteins, evolutionary relationship between species, biased mutation pressure, and gene expression (Muto and Osawa, 1987; Sémon et al, 2005). The lower G+C content of SARS-CoV-2 compared with SARS-CoV and MERS-CoV could affect the stability of its three-dimensional (3D) structure. Kudla et al. have showed that the G+C content of non-coding sequence in the promoter region (UTR) may affect the expression level of the gene. Kudla et al. found that there was a positive correlation between an excess of G+C content in the non-coding region and an increase in gene expression, and the efficiency of gene expression can increase by up to 100 times (Kudla et al., 2006). In this study, the 5′UTR G+C content was 44.6%, 43.5% and 44.7% for SARS-CoV-2, MERS-CoV and SARS-CoV, respectively. The difference of 1.1% between SARS-CoV-2 and MERS-CoV can lead to a faster and dramatic progression of many parameters related to viral infection, including the prognosis of the disease. The number of cases (76.8 million) and deaths (1.69 million) from its emergence support this approach.

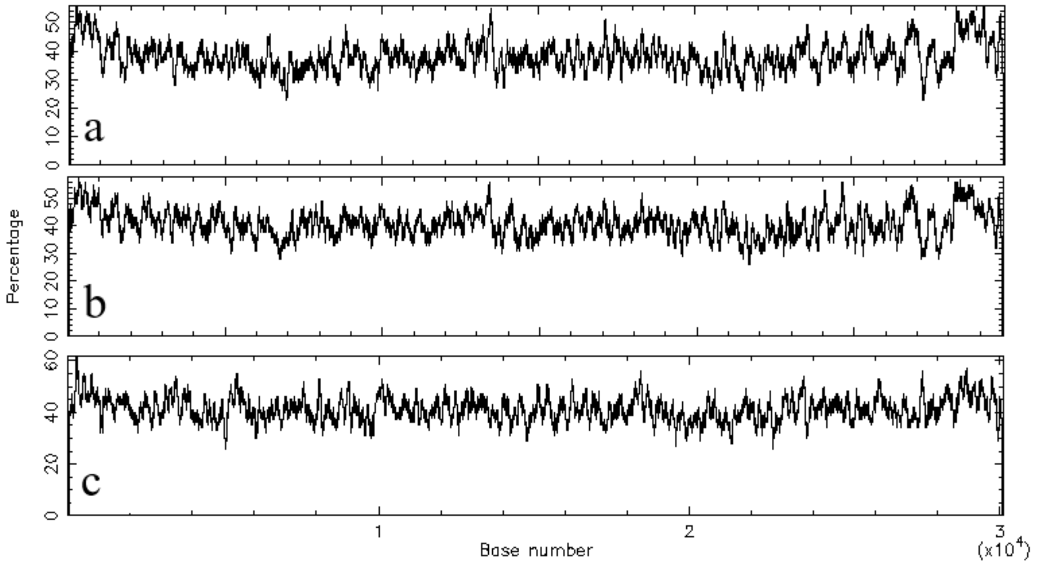

Figure 2. G+C content of SARS CoV-2 (a), SARS CoV (b) and MERS CoV (c)

In this study, SARS-CoV-2 and SARS-CoV showed 83.3% similarity, whereas SARS-CoV-2 and MERS-CoV showed 42.5% similarity at the amino acid level (Table 1). When the basic structural proteins were compared, the E, M, N and S proteins of SARS-CoV-2 and SARS-CoV were found to be 94%, 90.1%, 90.6% and 76.1% identical, respectively. In the comparison of SARS-CoV-2 and MERS-CoV, identity rates fell to 30%. The high similarity between SARS-CoV-2 and SARS-CoV in the linear sequence that determines the primer structure of the protein provides important hints about 3D structure and function of the protein (Panagiotou and Plaxco, 2020). Studies have shown that structural proteins of SARS-CoV-2 such as N, M and S, which can be targeted structures for vaccine and drug studies, show a similar homology with those of SARS-CoV (Ibrahim et al., 2020; Tilocca et al., 2020).

Table 1. Amino acid composition of SARS CoV-2, SARS CoV and MERS CoV

%	Ala A	Cys C	Asp D	Glu E	Phe F	Gly G	His H	Ile I	Lys K	Leu L
SARS CoV-2	6.83	3.07	5.10	4.81	5.00	5.94	1.87	5.15	5.92	9.65
SARS CoV	7.20	3.17	5.25	4.79	4.77	5.98	2.07	5.09	5.70	9.78
MERS CoV	7.33	3.01	5.31	3.87	5.05	5.84	2.06	4.79	5.26	9.47
%	**Met M**	**Asn N**	**Pro P**	**Gln Q**	**Arg R**	**Ser S**	**Thr T**	**Val V**	**Trp W**	**Tyr Y**
SARS CoV-2	2.21	5.41	3.94	3.65	3.40	6.75	7.51	8.14	1.11	4.54
SARS CoV	2.45	5.13	4.01	3.63	3.68	6.68	7.30	7.91	1.09	4.32
MERS CoV	2.26	4.99	4.21	3.50	3.63	7.67	6.94	8.96	1.14	4.71

In the present study, 11 physicochemical parameters were examined in amino acid sequences (Table 2). The ability of histidine (H) to respond to minute changes in the local pH value within the cell by changing its electric charge necessitates its presence in catalytic domains. Considering the genomic organisations, a similar percentage of H distribution between SARS-CoV and MERS-CoV and difference of 0.2% between SARS-CoV-2 and SARS-CoV can be associated with low genomic organisation of SARS-CoV-2. Disulphide bonds between cysteine (C) amino acids are directly related to the 3D structure, stability and therefore functionality of the protein. In this study, it was determined that C residues of SARS-CoV-2 and SARS-CoV were in positional identical at a ratio of 100%. As glycine (G) does not have a side chain, it is usually found within loop or coil sites. The G percentage densities and high levels of positional identity in the proteome indicate a high similarity among three viral organisms in terms of secondary and tertiary structures.

Table 2. Physicochemical properties of amino acids of three Coronaviruses

Property	Residues	Rate %		
		SARS CoV-2	SARS CoV	MERS CoV
Tiny	(A+C+G+S+T)	30.10	30.34	30.80
Small	(A+C+D+G+N+P+S+T+V)	52.70	52.63	54.27
Aliphatic	(A+I+L+V)	29.77	29.98	30.55
Aromatic	(F+H+W+Y)	12.52	12.25	12.95
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	55.58	55.76	56.77
Polar	(D+E+H+K+N+Q+R+S+T+Z)	44.42	44.24	43.23
Charged	(D+E+H+K+R+Z)	21.10	21.49	20.13
Basic	(H+K+R)	11.19	11.45	10.95
Acidic	(B+D+E+Z)	9.92	10.04	9.18
Isoelectric point		6.78	6.78	7.11

Molecular weight (kda)	1,580	1,595	1,593

Hydropathy analysis was performed using the Kyte–Doolittle scale (Figure 3). Polar and aromatic amino acids that determine superficial conformation of the protein, charged amino acids involved in the formation of important salt bridges for structural stability, high similarities and hydropathy values of hydrophobic amino acids that contribute to the protein core structure amino acids indicate high topological compatibility between SARS-CoV-2 and SARS-CoV. The study of Banerjee et al. on S protein supports our data and highlights the high topological similarity between SARS-CoV and SARS-CoV-2 in the N- and C-terminal domains (Banerjee et al., 2020).

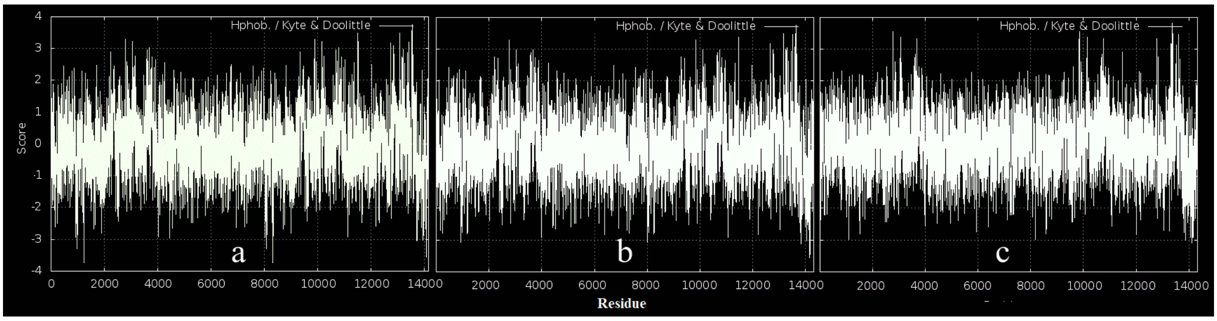

Figure 3. Hydropathicity value a) SARS CoV-2, b) SARS CoV, c) MERS CoV

RNA viruses have a high genetic mutation rate. This property leads to evolutionary differentiation and change in their virulence properties (Lin et al., 2019). Analysis of the probabilities of evolutionary change and the mutation risks of the nucleotides and amino acids for SARS-CoV-2 indicate targets with low mutation potential in regards to studies on drug and vaccine design and ensure that the durability of the treatment to be developed is long-term. In this study, the number of amino acid substitutions per site between sequences of SARS-CoV-2 and SARS-CoV was 0.183. The number of amino acid substitutions per site from between COVID19 and MERS sequences are 0.856. The extent of differences in base composition biases between COVID19, SARS and MERS sequences was statistically significant ($p<0.05$). Disparity Index per site is 0.329 and 1.783 for COVID19-SARS and COVID19-MERS sequences, respectively. Amino acid substitution pattern and mutation data matrices were estimated under the Jones-Taylor-Thornton model (Table 3). In the COVID19 protein structure, I (I> V and I> L), V (V> I and V> A), D (D> E and D> N), E (E> D) and R (R> K) amino acids are estimated to contain very high risk of mutation. The repeat numbers of 14 amino acids with high mutation risk in the COVID19 genome are shown in Table 4. The study of Huang et al. conducted with 125 samples and by analysing the mutations in the ORFs is in line with our data of mutation probabilities calculated for each residue (J.-M. Huang et al., 2020). Mutations in ORF1 at 1078T, 1219Y, 1574I, 1582P, 1808F, 2457K and 2517D positions indicate 14 residues with high mutation risk described in this study. Amino acid change was reported at 87 positions in total in the ORFs of SARS-

CoV-2. Similarly, the mutation data of the other ORFs were consistent with our mutation risk analysis, indicating that ORF6 and ORF7b were highly conserved.

Table 3. Amino acid substution and mutation data matrix*

From\To	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	-	0.14	0.12	0.22	0.06	0.12	0.34	0.67	0.03	0.10	0.15	0.11	0.06	0.03	0.51	1.37	1.39	0.01	0.02	1.01
R	0.21	-	0.10	0.04	0.11	0.64	0.10	0.53	0.38	0.07	0.18	2.01	0.05	0.01	0.19	0.35	0.20	0.09	0.04	0.06
N	0.22	0.12	-	1.48	0.03	0.16	0.19	0.30	0.48	0.13	0.06	0.78	0.04	0.02	0.03	1.79	0.71	0.00	0.12	0.06
D	0.33	0.04	1.22	-	0.01	0.11	2.49	0.49	0.12	0.03	0.03	0.09	0.02	0.01	0.03	0.21	0.13	0.00	0.08	0.11
C	0.23	0.27	0.07	0.03	-	0.02	0.02	0.21	0.09	0.04	0.08	0.02	0.05	0.14	0.03	0.76	0.14	0.08	0.35	0.21
Q	0.22	0.80	0.17	0.14	0.01	-	1.09	0.09	0.68	0.02	0.33	0.91	0.06	0.01	0.42	0.19	0.16	0.01	0.04	0.06
E	0.43	0.08	0.13	2.07	0.01	0.73	-	0.43	0.03	0.03	0.05	0.53	0.02	0.01	0.05	0.11	0.10	0.01	0.01	0.16
G	0.69	0.36	0.17	0.34	0.06	0.05	0.36	-	0.02	0.01	0.03	0.08	0.02	0.01	0.05	0.66	0.10	0.04	0.01	0.16
H	0.09	0.85	0.89	0.27	0.08	1.21	0.08	0.08	-	0.05	0.26	0.16	0.04	0.10	0.30	0.26	0.14	0.01	0.98	0.04
I	0.14	0.06	0.11	0.03	0.02	0.02	0.04	0.02	0.02	-	1.10	0.06	0.59	0.16	0.03	0.14	0.77	0.01	0.05	3.28
L	0.12	0.10	0.03	0.02	0.02	0.15	0.03	0.03	0.06	0.64	-	0.05	0.47	0.53	0.28	0.21	0.08	0.04	0.04	0.61
K	0.15	1.73	0.56	0.08	0.01	0.63	0.56	0.10	0.06	0.06	0.07	-	0.08	0.01	0.06	0.17	0.29	0.01	0.01	0.04
M	0.19	0.11	0.07	0.05	0.04	0.10	0.06	0.05	0.04	1.32	1.82	0.19	-	0.09	0.04	0.10	0.64	0.02	0.03	1.05
F	0.06	0.02	0.02	0.01	0.07	0.01	0.01	0.02	0.05	0.21	1.18	0.01	0.05	-	0.04	0.33	0.04	0.04	0.92	0.20
P	0.78	0.19	0.03	0.03	0.01	0.34	0.06	0.08	0.14	0.03	0.50	0.07	0.02	0.03	-	0.99	0.36	0.01	0.02	0.07
S	1.55	0.27	1.12	0.16	0.23	0.12	0.10	0.73	0.09	0.11	0.28	0.15	0.03	0.20	0.73	-	1.45	0.02	0.11	0.14
T	1.83	0.17	0.52	0.11	0.05	0.11	0.11	0.12	0.06	0.70	0.13	0.30	0.26	0.03	0.31	1.69	-	0.01	0.03	0.39
W	0.03	0.33	0.01	0.02	0.12	0.04	0.04	0.21	0.02	0.04	0.25	0.03	0.02	0.11	0.02	0.11	0.02	-	0.13	0.08
Y	0.06	0.06	0.15	0.12	0.22	0.05	0.02	0.02	0.70	0.08	0.11	0.03	0.02	1.15	0.03	0.22	0.06	0.06	-	0.06
V	1.16	0.05	0.04	0.08	0.07	0.04	0.15	0.18	0.01	2.60	0.83	0.04	0.37	0.12	0.06	0.14	0.35	0.02	0.03	-

*Substitution pattern and rates were estimated under the Jones-Taylor-Thornton (1992) model. Relative values of instantaneous r should be considered when evaluating them. For simplicity, sum of r values is made equal to 100. The amino acid frequencies are 7.69% (A), 5.11% (R), 4.25% (N), 5.13% (D), 2.03% (C), 4.11% (Q), 6.18% (E), 7.47% (G), 2.30% (H), 5.26% (I), 9.11% (L), 5.95% (K), 2.34% (M), 4.05% (F), 5.05% (P), 6.82% (S), 5.85% (T), 1.43% (W), 3.23% (Y), and 6.64% (V). This analysis were conducted MegaX.

In conclusion, SARS-CoV-2 was shown to be highly structurally and functionally similar to another type of betacoronavirus, SARS-CoV. For SARS-CoV-2, residues with high mutation risk and their repeat numbers and general location information in the genome were described. It is predicted that ORF1, which contains a large number of residues with high mutation risk and encodes the replicas of SARS-CoV-2, and S proteins involved in binding to the host cell receptor may be risky targets for drug and vaccine studies. It is believed that the consideration of highly conserved non-structural proteins such as ORF6 and ORF7b as target structures will contribute to the durability of the treatments to be developed.

Table 4. Amounts of residues with high mutation potential in SARS CoV-2 proteins

From/to	r	PR	ORF1b	ORF1a	Spike	ORF3a	E	M	ORF6	ORF7a	ORF7b	ORF8	N	ORF10
I > V	3.28	**I**	343	215	76	21	3	20	10	8	5	10	14	3
I > L	1.10													
V > I	2.60	**V**	598	371	97	25	13	12	3	8	1	12	8	4
V > A	1.16													
D > E	2.49	**D**	389	211	62	13	1	6	4	2	2	7	24	1
D > N	1.22													
E > D	2.07	**E**	340	239	48	11	2	7	5	8	3	6	12	0
R > K	2.01	**R**	244	131	42	6	3	14	1	5	0	4	29	2
T > A	1.83	**T**	527	345	97	24	4	13	3	10	1	5	32	2
T > S	1.69													
M > L	1.82	**M**	168	105	14	4	1	4	3	1	2	1	7	2
M > I	1.32													
M > V	1.05													
N > S	1.79	**N**	384	233	88	8	5	11	4	2	1	2	22	5
N > D	1.48													
K > R	1.73	**K**	434	276	61	11	2	7	4	7	0	5	31	0
S > A	1.55	**S**	456	294	99	22	8	15	4	7	2	9	37	2
S > T	1.45													
S > N	1.12													
A > T	1.39	**A**	487	309	79	13	4	19	1	9	2	5	37	2
A > S	1.37													
A > V	1.01													
F > L	1.18	**F**	349	208	77	14	5	11	3	10	6	8	13	4
Y > F	1.15	**Y**	335	195	54	17	4	9	2	5	1	7	11	3
Q > E	1.09	**Q**	239	151	62	9	0	4	3	5	1	6	35	1

Abbrevations. E: Envelope M: Membrane N: Nucleocapsid phosphoprotein ORF: Open reading frame PR: Residue with mutation potential

References

Abubucker, S., Martin, J., Taylor, C. M. and Mitreva, M. 2011. "HelmCoP: An online resource for Helminth functional genomics and drug and vaccine targets prioritization". PLoS One, 6(7), 1-12.

Al-Osail, A. M. and Al-Wazzah, M. J. 2017. "The history and epidemiology of Middle East respiratory syndrome corona virüs". Multidisciplinary Respiratory Medicine, 12(20), 1-6.

Badani, H., Garry, R. F. and Wimley, W. C. 2014. "Peptide entry inhibitors of enveloped viruses: The importance of interfacial hydrophobicity". Biochimica et Biophysica Acta (BBA) – Biomembranes, 1838(9), 2180–2197.

Baltimore, D. 1971. "Expression of animal virus genomes". Bacteriological Reviews, 35(3), 235–241.

Banerjee, A., Santra, D. and Maiti, S. 2020. "Energetics based epitope screening in SARS

CoV-2 (COVID 19) spike glycoprotein by Immuno-informatic analysis aiming to a suitable vaccine development”. bioRxiv, 021725, 1-28.

Carroll, H., Beckstead, W., O’Connor, T., Ebbert, M., Clement, M., Snell, Q. and Mcclellan, D. 2007. “DNA reference alignment benchmarks based on tertiary structure of encoded proteins”. Bioinformatics, 23(19), 2648–2649.

Chan, J. F. W., Yuan, S., Kok, K. H., To, K. K. W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C. Y., Poon, R. W. S., Tsoi, H. W., Lo, S. K. F., Chan, K.H., Poon, V. K. M., Chan, W. M., Ip, J. D., Cai, J. P., Cheng, V. C. C., Chen, H., Hui, C. K. M. and Yuen, K. Y. 2020. “A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster”. Lancet, 395(10223), 514–523.

Chatterjee, S. 2020. “Understanding the nature of variations in structural sequences coding for Coronavirus spike, envelope, membrane and nucleocapsid proteins of SARS CoV-2”. SSRN Electron Journal, 1-12.

Drosten, C., Günther, S., Preiser, W., Van der Werf, S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A. M., Berger, A., Burguière, A. M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J. C., Müller, S., Rickerts, V., Stürmer, M., Vieth, S., Klenk, H. D., Osterhaus, A. D. M. E., Schmitz, H. and Doerr, H. W. 2003. “Identification of a novel coronavirus in patients with severe acute respiratory syndrome”. New England Journal of Medicine, 348(20), 1967–1976.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J. and Cao, B. 2020a. “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China”. Lancet, 395(10223), 497–506.

Huang, J. M., Jan, S. S., Wei, X., Wan, Y. and Ouyang, S. 2020b. “Evidence of the Recombinant Origin and Ongoing Mutations in Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)”. bioRxiv, 993816, 1-16.

Ibrahim, I. M., Abdelmalek, D. H., Elshahat, M. E. and Elfiky, A. A. 2020. “COVID-19 spike-host cell receptor GRP78 binding site prediction”. Journal of Infection, 80, 554–562.

Jones, D. T., Taylor, W. R. and Thornton, J. M. 1992. “The rapid generation of mutation data matrices from protein sequences”. Bioinformatics, 8(3), 275–282.

Katoh, K. 2002. “MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform”. Nucleic Acids Research, 30(14), 3059–3066.

Katoh, K., Rozewicki, J., Yamada, K. D. 2018. “MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization”. Briefings in Bioinformatics, 20(4), 1160–1166.

Kudla, G., Lipinski, L., Caffin, F., Helwak, A. and Zylicz, M. 2006. “High guanine and cytosine content increases mRNA levels in mammalian cells”. PLoS Biology, 4(6), 933–942.

Kumar, S. and Gadagkar, S. R. 2001. “Disparity index: A simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences”. Genetics, 158, 1321–1327.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. 2018. “MEGA X: Molecular evolutionary genetics analysis across computing platforms”. Molecular Biology and Evolution, 35(6), 1547–1549.

Lam, T. T. Y., Shum, M. H. H., Zhu, H. C., Tong, Y. G., Ni, X. B., Liao, Y. S., Wei, W., Cheung, W. Y. M., Li, W. J., Li, L. F., Leung, G. M., Holmes, E. C., Hu, Y. L. and Guan, Y. 2020. “Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China”. bioRxiv, 945485, 1-22.

Li, C., Yang, Y. and Ren, L. 2020. “Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species”. Infection Genetics and Evolution, 82(104285), 1-3.

Lin, R. W., Chen, G. W., Sung, H. H., Lin, R. J., Yen, L. C., Tseng, Y. L., Chang, Y. K., Lien, S. P., Shih, S. R. and Liao, C. L. 2019. “Naturally occurring mutations in PB1 affect influenza A virus replication fidelity, virulence, and adaptability”. Journal of Biomedical Science, 26(55), 1-14.

Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., Smoot, J., Gregg, A. C., Daniels, A. D., Jervey, S. and Albaiu, D. 2020. “Research and development on therapeutic agents and vaccines for COVID-19 and related human Coronavirus diseases”. ACS Central Science, 6(3), 315–331.

Mahajan, M., Chatterjee, D., Bhuvaneswari, K., Pillay, S. and Bhattacharjya, S. 2018. “NMR structure and localization of a large fragment of the SARS-CoV fusion protein: Implications in viral cell fusion”. Biochimica et Biophysica Acta-Biomembranes, 1860(2), 407–415.

McBride, R. and Fielding, B. C. 2012. “The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis”. Viruses, 4(11), 2902–2923.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. and Lopez, R. 2013. “Analysis tool web services from the EMBL-EBI”. Nucleic Acids Research, 41(1), 597–600.

Muto, A. and Osawa, S. 1987. “The guanine and cytosine content of genomic DNA and bacterial evolution”. Proceedings of the National Academy of Sciences of the USA, 84(1), 166–169.

Overington, J. P., Al-Lazikani, B. and Hopkins, A. L. 2006. “How many drug targets are there?”. Nature Reviews Drug Discovery, 5(12), 993–996.

Panagiotou, E. and Plaxco, K. 2020. “A topological study of protein folding kinetics”. Topology and Geometry Biopolymers, 746, 223–234.

Pardi, N., Hogan, M. J., Porter, F. W. and Weissman, D. 2018. “mRNA vaccines-a new era in vaccinology”. Natere Reviews Drug Discovery, 17(4), 261–279.

Regla-Nava, J. A., Nieto-Torres, J. L., Jimenez-Guardeño, J. M., Fernandez-Delgado, R., Fett, C., Castaño-Rodríguez, C., Perlman, S., Enjuanes, L. and DeDiego, M. L. 2015. “Severe acute respiratory syndrome Coronaviruses with mutations in the E protein are attenuated and promising vaccine candidates”. Journal of Virology, 89(7), 3870–3887.

Rice, P., Longden, L. and Bleasby, A. 2000. “EMBOSS: The European Molecular Biology Open Software Suite”. Trends in Genetics, 16(6), 276–277.

Sémon, M., Mouchiroud, D. and Duret, L. 2005. “Relationship between gene expression and GC-content in mammals: Statistical significance and biological relevance”. Human Molecular Genetics, 14(3), 421–427.

Shen, S., Wen, Z. L. and Liu, D. X. 2003. “Emergence of a coronavirus infectious bronchitis virus mutant with a truncated 3b gene: Functional characterization of the 3b protein in pathogenesis and replication”. Virology, 311(1), 16–27.

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N. and Siddique, R. 2020. “COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses”. Journal of Advanced Research, 24, 91–98.

Tilocca, B., Soggiu, A., Sanguinetti, M., Musella, V., Britti, D., Bonizzi, L., Urbani, A. and Roncada, P. 2020. “Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses”. Microbes and Infection, 22(4), 188-194.

Ul-Qamar, M. T., Alqahtani, S. M., Alamri, M. A., Chen, L. L. 2020. “Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants”. Journal of Pharmaceutical Analysis, 1-7.

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T. and Veesler, D. 2020. “Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein”. Cell, 181(2), 281-292.

Weiss, S. R. and Leibowitz, J. L. 2011. “Coronavirus pathogenesis”. Advances in Virus Research, 81, 85–164.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. and Hochstrasser, D. F. 1999. “Protein identification and analysis tools in the ExPASy server”. 2-D Proteome Analysis Protocols, 112, 531–552.

Worldometer. “Coronavirus Cases”, https://www.worldometers.info/coronavirus/? , 21.12.2020

Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. and Fouchier, R. A. M. 2012. “Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia”. New England Journal of Medicine, 367(19), 1814–1820.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F. and Tan, W. 2020. “A novel coronavirus from patients with pneumonia in China 2019”. New England Journal of Medicine, 382, 727–733.

Zuckerkandl, E. and Pauling, L. 1965. “Evolutionary divergence and convergence in proteins”. Evolving Genes and Proteins, *Academic Press*, pp. 97–166.