Iranian Journal of Virology 2021;15 (1): 7-14

Original Article

A Novel Genetic classification of SARS coronavirus-2 following whole nucleic acid and protein alignment of the isolated viruses

Sanchooli A[1], Shahkarami MK[1], Thekkiniath J[2], Karimi Naghlani S[1], Kamali M[3], Cheraghi M[1], Shojaei M[1*]

1. Razi Vaccine and Serum Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran
2. Fuller Laboratories, 1312 East Valencia Drive, Fullerton, CA- 92831, USA
3. Department of Biology, Islamic Azad University, North Tehran Branch, Tehran, Iran

Abstract

Background and aims: The end of 2019 has marked the year, which the human population encountered a novel virus; SARS-CoV-2 that causes a disease namely COVID-19. Here we focused on the genome and protein mutations and subsequently suggested a new classification of the SARS-CoV-2.

Materials and Methods: Our study showed that some extra positions in the virus genome play a key role in the SARS-CoV-2 classification. Based on the analysis of the whole genome sequences of 93 viruses.

Results: mutations were classified into nine divisions including IA-1, IA-2, IA-3, IB, II, L1, L2, L3 and S. Totally, 279 mutations were found in the SARS-CoV-2 genomes. 24 mutations lead to the amino acid frame shifting, of which 15 mutations lead to positive frame shifting in amino acids sequences.

Conclusion: Sequence alignment of these positions with that of ancestors showed no change suggesting that they might have occurred in the SARS-CoV-2 genomes to adapt itself to humans.

Keywords: SARS-coronavirus-2; COVID-19; classification; Phylogenetic tree analysis; Mutations; Evolution

Introduction

During last two decades, three viruses including severe acute respiratory syndrome coronavirus (SARS-CoV), middle east respiratory syndrome (MERS-CoV), and the novel coronavirus (SARS-CoV-2) belong to Coronaviridae family have breached the barrier, crossing over to interspecies, and have selected humans as their terminal host. Of these three viruses, SARS-CoV-2 has caused the highest morbidity (1). SARSCoV-2 or COVID-19 is more contagious than SARS-CoV and MERS-CoV (2).

[*]**Corresponding author:**
Mohammad Shojaei.
Email: Shojaei.mohamaddr@yahoo.com.

Until now, according to worldo-meters website globally it has caused more than 33,062,174 cases and 998,803 deaths as of September 27, 2020.

SARS-CoV-2 is an enveloped positive-sense single-stranded RNA virus with 29.87 Kb genome and 37-38% GC content (3). Due to its structural similarities with the SARS-CoV, it has been named as SARS-CoV-2 (4, 5). Based on phylogenetic tree analysis, Sarbecoviruses have been classified into three clades. SARS-CoV-related strains formed clade 1, which includes SARS-CoV-2 and bat-SL-CoVZC4.

The bat-SL-CoVZXC21 formed clade 2, SARS-CoV strains from humans and many genetically similar SARS-like coronaviruses from bats formed clade 3 SARS-CoV-2, RaTG13, and Bat-SARS-like coronaviruses formed a single cluster in phylogenetic tree

analysis (6, 7). The whole-genome sequence of SARS-CoV-2 indicates discordant clustering with Bat-SARS-like coronavirus, particularly in the first 11,498 nucleotides in the 5' region and the last 24,341-30,696 nucleotides in the 3' region of the genome (7). Also, Lack of a clear relationship of SARS-CoV-2 to its related Sarbecoviruses suggests that recombination has less effect on the emergence of this virus. Thus, it appears that this virus was produced by mutations that occurred under natural selection. (1, 7). But, more than 99.5% identity of its genomic sequences from patients sugges-ting that this virus has selected human species as a terminal host recently (8).

SARS-CoV-2 is 88% similar to two bat-derived severe acute respiratory syndrome (SARS)-like coronaviruses, bat-SL-CoVZC45, and bat-SL-CoVZXC21. It has nucleotide similarity with bat coronavirus strain bat-SL-CoVZC45 and bat-SL-CoVZXC21 based on ORF1a/1b, S, and N genes. It has about 79% similarity with SARS coronavirus and 50% similarity with MERS coronavirus (6). The similarity of SARS-CoV-2 with a bat SARS-related coronavirus (RaTG13) isolated in 2013 coronavirus is 96% (less than 4% difference) (8). It was found that SARS-CoV-2 and bat SARS-CoVs have the same amino acid codon usage, whereas two genes have rather distinct synonymous codon usage patterns (9). SARS-CoV-2 may have two ancestors (SARS-like bat viruses: bat-SL-CoVZC45, bat-SL-CoVZXC 21) (10). The genome of SARS-CoV-2 includes the structural proteins including Membrane (M), Spike (S), Envelope (E) and Nucleoprotein (N) as well as some non-structural and accessory proteins (ORF3, ORF7, ORF8, ORF 9, ORF 10b, ORF 13, ORF 14 proteins) (3, 11). Based on the accessory proteins, the source of SARS-CoV-2 could be bat coronaviruses, because the ORF3 and ORF8 accessory proteins and the major charact-eristics of bat coronaviruses are also found to be present in SARS-CoV-2.

Additionally, phylogenetic tree analysis showed that SARS-CoV-2 is related to bat coronaviruses. Thus, SARS-CoV-2 and bat coronaviruses have more similarities than other coronaviruses (3, 12). But, transmission to humans has occurred via an intermediate host. Understanding of genomics and proteomics of SARS-CoV-2 may provide insights for developing drugs and vaccines to combat COVID-19. Here, we provide detailed geno-mics and evolution of SARS-CoV-2 (13).

Based on the evolutionary rate for SARS-CoV-2 and 99% identity between its sequences derived from different patients, it was indicated that this pandemic has one source (6, 14). It has been noted that subgenus Sarbecovirus has several recombinations (15). Some researchers believe that SARS-CoV-2 might have created by the recombination of a pangolin coronavirus with a Bat-CoV-RaTG13 virus (16). Genomic analysis of bat coronaviruses ZC45 and ZCS21 showed that they have a recombinant genome containing several fragments of SARS-related coronaviruses. SARS-CoV-2 which belongs to this lineage has a recombinant fragment of pangolin coronaviruses (15).

These data showed that recombination has occurred in the S protein of Pangolin-CoV. The region of nucleotides 1-914 in upstream of the genome was similar to Bat SARS-CoV ZXC21 and Bat SARS-CoV ZC45, while the remaining part of this gene is similar to SARS-CoV-2 and Bat-CoV-RaTG13 (16).

The amino acid sequence of SARS-CoV-2 differs from other coronaviruses in the ORF1ab and S proteins (17). The recent ancestor of this virus accumulated several amino acid substitutions in the receptor binding domain (RBD) of S protein and in ORF1a polyprotein that is critical for replication and transcription. It was carried on the scenario in the emergence of the SARS-CoV-2, which has been done by some underwent convergent evolution recombinants. Recombination and rapid evolution in the nonstructural protein 3 (nsp3) and Spike genes of the bat, civet, and human SARS coronaviruses resulted in the emergence of SARS-CoV-2 (18, 19). The majority of the SARS-CoV-2 proteins are homologous (95%–100%) with its counterpart in SARS-CoV.

This suggests that it may cause the same evolutionary line (20). This evolution mostly was done by mutations. Several studies showed

that SARS-CoV-2 had some mutations during spreading among humans.
By analyzing 93 whole genome sequences, we studied an additional classification of this virus according to the whole genome sequences similarity.

Methods

Providing the phylogenetic tree analysis: The GenBank was searched for the detection of full genome sequences. 93 of the full genome were extracted from the NCBI website and saved into notepad software. Then it was submitted to Clustal Omega online software to be aligned. The Jalview version of the alignment was downloaded from the result of Clustal Omega online software. The data was changed to the FASTA format by jalview software. Data was opened with MEGA-X software and check for alignment. The evolutionary history in MEGA-X software was inferred using the Neighbor-Joining method. The optimal phylogenetic tree with the sum of the branch was estimated using the RAXMIGUI (21). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The new data was opened by Recombination detection program (RDP version4). A whole-genome nucleotide sequence of SARS-CoV-2 was extracted from GenBank and opened by SnapGene software. Every mutation was checked in RDP software. Then the mutation was found in SnapGene software for reporting the amino acid codons which were changed. The whole-genome from nucleotide position 50 to 29840 was checked and its nucleotide codon and amino acid before and after changing were reported in a table. High variation positions were checked for a relationship with them by typing of the virus. Finally, the new classification table was provided. Then classification was checked in phylogenetic tree analysis and found a new branch of this virus, which was not recorded before.

Results

By searching the GenBank, it was indicated that till study date 93 whole-genome sequences have existed. Thus, they were extracted and changed its format to FASTA format. The submission to Clustal Omega online software was done successfully and a Jalview format was downloaded as the result. Using the MEGA-X software showed an evolutionary history by the Neighbor-Joining method. Thus a phylogenetic tree analysis was drawn by RAXMIGUI software. The file was successfully analyzed by Recombination detection program (RDP version4). Detection of all mutations was successfully done by the RDP program. Mutations were detected between position nucleotide 50 to position nucleotide 29840. 279 mutations were found in the whole genome of SARS-CoV-2 during this research. Out of 279, 24 mutations resulted in a change in the amino acid sequences. But, 70 mutations had no change in amino acid sequences. Thus, the human codon usage was done to understand 15 positions faced to positive changing, 24 positions faced to negative changing, while 4 positions have no changes in amino acid sequences or codon usage fractions.
As a matter of fact, 8, 111, 16, 4, 2, 2, 2, 21, 1, and 4 nucleotide positions met the mutations in 5UTR, Orf1ab, S, Orf3, E, M, Orf8, Orf10 and, 3UTR respectively.
In addition to this data analyzing by RDP software, it was shown that some nucleotide positions had more than one mutation. Thus, they helped to change the classification category of this virus. Six positions had high mutations, suggesting that SARS-CoV-2 tend to be divided into some subdivisions .
Phylogenetic tree analysis of viruses admitted the new classification of this virus into subdivisions. IA was divided into IA-1, IA-2, IA-3 and L was divided into L1, L2 and L3 (Table 4.).

Table 1. Completion of typing of SARS CoV-2 based on more mutations. As it is indicated, 7 positions play critical role on SARS CoV-2 typing of which 4 positions were not reported before. P.: position. Prevalence was estimated according to 93 whole genome sequences analyzed in this study.

		S					II	L	
		IA			IB				
		IA-1	IA-2	IA-3			L1	L2	L3
Variations	4410	T	C	T	T	C/T	T	T	
	5070	G	T	G	G	T/G	G	G	
	8790	T			C	C/T	C/T	C	
	17381	C			C	C	C	C	T
	17755	C		T	C	C/T	C	C	
	17866	A		G	A	G/A	A	A	
	28155	C			C	C	T	T	
	29106	T			C	C/T	C	C	
	18068	T			C	C/T	T	C	
	26155	G			G	G	G	T	G
Prevalence		8.6	4.3	6.4	4.3	1.1	0	64.5	7.5	4.3

Table 2. Novel classification is suggested by analyzing the 93 whole genomes of SARS-CoV-2 *The name of sequences were used according to the references of the table 3

Classified name		Name of sequences*
IA	IA-1	66,67,71,74,75,81,83-85
	IA-2	61, 68-70
	IA-3	72, 73, 76-78,80
IB		64, 65, 82,79
II		-
L	L1	1- 3, 6-13, 15-41, 43-46, 49-56,59-60,62,86- 93
	L2	4, 5, 92, 42, 48, 90, 57
	L3	38, 47, 63, 14
S		61,64- 85

Discussion

COVID-19 virus has two major types of mutations including L type and S type. The S type is the ancestor of L type due to finding that in pangolins coronaviruses and bat SARS-like coronaviruses .

The L type has appeared under the evolution of the S type. These types are diagnosed by two single nucleotide polymorphisms (SNPs including SNP1 at position 8,782 in orf1ab and SNP2 at position 28,144 in ORF8. SNP1 changes the codon AGT (Ser) to AGC (Ser).

Thus, the SNP1 is a synonymous mutation but induced more expression ATF6 (Activating transcription factor 6), an endoplasmic reticulum (ER)-localized protein, from ORF8 in human cells (22). SNP2 changes the codon TCA (Ser) to TTA (Leu), thus SNP2 is non-synonymous (8) The L type is more prevalent than S type with 70% and 30% prevalence, respectively (Table 1). Additionally, some studies have suggested that L type is more aggressive than S type (8, 23).

Phylogenic tree analysis revealed that SARS-CoV-2 has two clades: type I and Type II .(^)

Type II is more prevalent and appeared due to the evolution of type I. while Type I has been reported in the Bat CoV RaTG13. It seems that the outbreak of type II COVID-19 probably occurred in the Huanan market was highly contagious than type I. Additionally, type II is

more efficient than type I in terms of protein expression. All the positions in the nucleotide sequence occurred by changing T to C nucleotide or vice versa. In fact, type I and Type II differ in three positions at 8750 (synonymous mutation in ORF1ab), 28112 (nonsynonymous mutation leading to a change from Leucine to Serine in the ORF8 gene), and 29063 (synonymous mutation in N gene).

Based on the last position (29063), type I is divided into two subtypes, IA and IB. Of these, IA is the ancestor of type II and is the earliest transmission source. Both synonymous mutations cause more efficient protein expression in type II but not in type I. Thus, type I outbreak was occurred in another place, but not in the Wuhan market, while type II occurred in the Huanan market. A recent study suggested that patients infected with two types of this virus (mentioned above) need different treatments due to difference in the acceleration of illness onset and efficient translation (10, 24) (Figure 1 and Table 1) .

By analyzing the diversity of the SARS-CoV-2 full genome, we found more positions for the classification of this virus than it has been reported before. This is the first study that reports that in addition to the 3 positions including 8790, 28155, 29106 (10), six more positions including 4410, 5070, 17381, 17755, 17866, and 26155 also important in classification of this virus. The phylogenetic tree of the whole-genome analysis showed some new branches. It was indicated that the IA and L types were divided into 3 subdivisions. These divergences may be associated with a variation of three positions 4410, 5070, and 17866 (Figure 1). Sequence alignment of these positions with that of ancestors showed no change suggesting that they might have occurred in the SARS-CoV-2 genomes to adapt itself for humans. One of the justifications to this extra adaptation is that position 4410 is a codon of amino acid leucine and by checking the human codon usage, it was indicated that changing the codon improved the protein expression of amino acid leucine from 0.13 to 0.20 fraction (Table 1). It can be classified as a subtype of IA as IA-2 which has variations in 4410 and 5070 positions within type IA. Another subtype would be IA-3 which has variations in 17755 and 17866 positions. Geographically, the prevalence of these subtypes was IA-2 in china and IA-3 in the USA. Also, according to variations in positions 26155 and 17381, L type can be divided into three subtypes L1, L2, and L3 (Table 1).

Moreover, it was revealed in our study all of IA divisions and IB were covered by the S type.

Analysis of 93 whole-genome sequences from nucleotide position 50 to the nucleotide position 29840 of SARS-CoV-2 showed that 21.5% (20) viruses had no mutations, while 279 mutations had happened in 78.5% (73) viruses. Thus, there were approximately three mutations were occurred per virus.

Interestingly, 7.88% (22) of whole mutations caused to appear the frameshifting in amino acid sequences. Additionally, 8.6% (24) out of 264 mutations resulted in the exchange of more than one amino acid. Viruses with more than six mutations are 91 with 11 mutations, 66 with 11 mutations, 77 with 10 mutations, 78 with 8 mutations, 62 with 8 mutations, 82 with 7 mutations, 87 with 7 mutations, 89 with 7 mutations, 71 with 7 mutations. All of these mutations were found to be from the USA.

Also, about 25.8% (24) of all mutations involved more than three nucleotides. Interestingly, more than 25.08% (70) of whole mutations did not result in the exchange of amino acids. However, they caused positive or negative effects on the translation process (according to human codon usage table). Thus, they may affect the severity of disease of which 7.16% (20) out of 279 mutations were positive and 23.29% (65) were negative mutations .

While 1.79% (5) had no effects on the translation process, 24 mutations caused it to produce a frameshifting in amino acid sequences. However, 70 different mutations had no change in amino acid sequences. Thus, by comparing with the human codon usage we found that 15 positions had positive mutations, 24 positions had negative mutations, while 4 positions had no changes in amino acid sequences or codon usage fractions. Our analysis showed that 8, 111, 16, 4, 2, 2, 2, 21,

1, and 4 nucleotide positions met the mutations in 5'UTR, Orf1ab, S, Orf3, E, M, Orf8, Orf10 ,and 3'UTR respectively. Additionally, 2.86% (8) occurred in upstream of the genome in 3'UTR, 63.44% (177) in Orf1ab, 7.16% (20) in S gene, 3.58% (10) in Orf3a gene, 1.07% (3) in E gene, 17.92% (50) in Orf8 gene, 1.13% (3) in Orf10 gene and more than 1.7% (5) of them happened in downstream of the genome 5'UTR (Table2,3). Some of these genes had more evolution than others suggesting adaptation of this virus to choose human as a final host.

Conclusion

Based on genome analysis and comparison of the SARS-CoV-2 with their ancestors and related viruses, we predict that this virus has a mutation potential. So, genomic analysis is essential for the determining the level of aggression and it will help the treatment and prevention of this disease.

In addition to three nucleotide positions which were reported previously, in this study we found six extra other positions including 4410, 5070, 17381, 17755, 17866, and 26155 playing a key role in classification of this virus.

Sequence alignment of these positions with that of ancestors showed no change suggesting that they might have occurred in the SARS-CoV-2 genomes to adapt itself to humans. So, 6 new groups may be added to the classification of this virus as the subdivisions. Thus, IA-1, IA-2 and 3, groups would be divided as subdivisions of IA type of this virus with the prevalence of 8.6, 4.3 and 6.4 percent, respectively. Of these subtypes, while IA-1 spread across the world, IA-2 and IA-3 could be found in china and USA, respectively.

Meanwhile all of these subdivisions belong to S type. According to variations of L type in positions 26155 and 17381 of nucleotide sequences, it would be divided into three subdivisions L1, L2 and L3 with the prevalence of 64.4, 7.5 and 4.3 percent, respectively.

Our evaluation of the genome of SARS-CoV-2 suggests that 7.88% of mutations resulted in the frame-shift of amino acid sequences and 8.6% of mutations had more than one amino acid variation. Furthermore, the highest mutations were reported in the USA. While 25.08% of mutations did not result in exchanging the amino acid sequences. Others caused positive or negative effects on protein translation.

Thus, it seems that they has affect the severity of their disease. We found that 7.16% of mutations were positive and 23.29% were negative mutations. Also, 63.44, 7.16, 3.58, 1.07, 17.92 and 1.13 percent were occurred in Orf1ab, S, Orf3a, E, Orf8, Orf10 genes, respectively. Further investigation on the breakdown of species barrier between animals and humans of this virus will help us to prevent new zoonosis diseases from other bat-CoVs. More studies are critical to understand the genetics and biology of this virus to prevent and treat this disease

Acknowledgment

Consideration of these topics benefited from discussion and papers published with many individuals. Special thanks to all the scientists who tried to improve our knowledge about genome and the scientists, we used their SARS CoV-2 sequences had submitted in NCBI.

Conflict of interest

The authors declare that there is no conflict of interest.

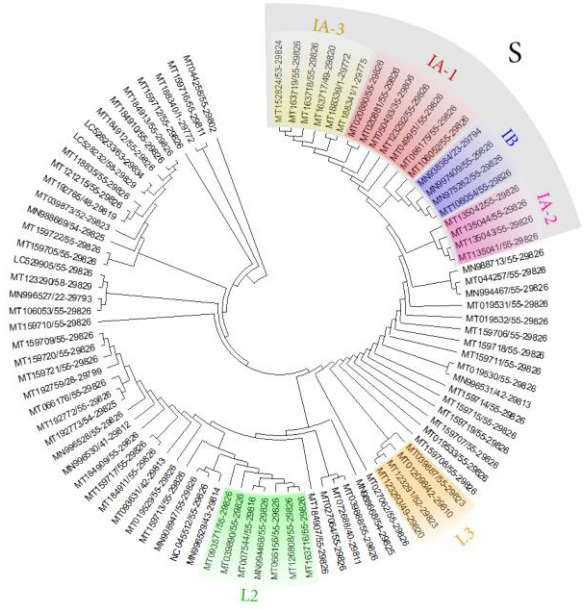
Funding

This research was funded by Razi vaccine and serum research institute

Table 3. Genetic diversity of SARS coronavirus 2 according to mutations which were found in 93 sequences. FS: frame shift, fractions of amino acids was done according to human codon usage, sequence numbers (Seq. number) was measured according to alignment number.

Seq. number	Ref.	Mutation	Result	Seq. number	Ref.	Mutation	Result	Seq. number	Ref.	Mutation	Result
5'UTR											
83	(67)	C→A	-	112	(55)	T→A	-	119-120	(55)	TT→CG	-
127-128	(55)	CT→GC	-	132	(55)	G→A	-	194	(15)	C→T	-
249	(6)	C→T	-	262	(87)	C→T	-				
Orf1ab											
498	(66)	GAT→GAA	D→E	498	(62)	GAT→GA-	FS	515-530	(45,25)	TGGTCATGTTA TGGT→-	FS
515	(3)	CAT→CAC	H(0.41→0.51)	622	(51)	GCT→ACT	A→T	662	(63)	GGA→GAA	G→E
693-702	(45)	AAGTCATTT →-	K,S,F del.	1071	(91)	TCA→T-A	FS	1110	(78)	TCC→TCT	S(0.22→0.18)
1393	(21)	CAC→TAC	H→Y	1556	(71)	AGC→AAC	S→N	1699	(74)	ATT→GTT	I→V
1919	(87)	TCC→T-C	FS	2099	(50)	ACT→ATT	T→I	2277	(90)	GCA→GCT	A (0.23→0.26)
2285	(38)	ATT→ACT	I →T	2454	(92)	ACT→ACC	T(0.24→0.36)	2725	(42)	GGT→AGT	G→S
2979	(48)	ATG→ATT	M→I	3045	(6)	TTC→TTT	F (0.55→0.45)	3107	(24,91)	ACT→ATT	T→I
3185	(66)	CCT→CTT	P→L	3185	(62)	CCT→C-T	FS	3267	(34)	CAG→CAT	Q→H
3419	(92)	GCT→GTT	A→V	3526	(46)	GTT→TTT	V→F	3746	(36)	CCT→CTT	P→L
3786	(56)	ACA→ACG	T (0.28→0.12)	4410	(61, 68-70)	CTT→CTC	L (0.13→0.20)	5070	(61, 68-70)	TTG→TTT	L→F
5092	(51)	ATA→GTA	I→V	5488	(46)	GCC→-CC	FS	5580	(92)	ATG→ATT	M→I
5792	(72)	ACT→ATT	T→I	5853	(26)	AAA→AAT	K→N	6034	(37)	CCA→TCA	P→S
6039	(48)	AAC→AAT	N (0.54-0.46)	6044	(77)	AGC→GGC	S→G	6509	(74)	CCA→CTA	P→L
6644	(29)	ACT→ATT	T→I	6703	(38)	CCT→TCT	p→S	6827	(14,63)	AGT→ATT	S→I
7004	(54, 63)	ATC→ACC	I→T	7024	(9)	GGT→AGT	G→S	7874	(52)	GGT→GTT	G→V
8009	(7)	GAT→GCT	D→A	8396	(56)	AAC→AGC	N→S	8790	(61,64-85)	AGC→AGT	S (0.24→ 0.15)
8790	(62)	AGC→AG-	FS	8995	(56)	TTT→ATT	F→I	9042	(44)	AAA→AAG	K (0.42→ 0.58)
9165	(87)	TTT→TTC	F (0.45→ 0.55)	9282	(42)	AGA→AGG	R (0.20→ 0.20)	9482	(35)	GCT→GTT	A→V
9499	(44)	CAT→TAT	H→Y	9542	(7)	ACT→ATT	T→I	9569	(64)	TCA→TTA	S→L
9932	(40)	GCA→GTA	A→V	10044	(33)	ATC→ATT	I (0.48→0.36)	10091	(91)	TCT→T-T	FS
10240	(1,2)	CGT→TGT	R→C	10240	(2)	ATT→ACT	S→L	10515	(91)	AAC→AAT	N (0.54→0.46)
11091	1,65,86-89,91-9	TTG→TTT	L→ F	11091	(67)	TTG→TTC	L→F	11091	(90)	TTG→TT-	FS
11418	(19,36)	TTG→TTA	L→L	11758	(29)	CTC→TTC	L→F	11964	(29)	GAC→GAT	D (0.54→0.46)
12123	(48)	TCC→TCT	S (0.22→0.18)	12123	(48)	ATT→ACT	I→T	12481	(37)	CTA→TTA	L (0.07→0.07)
12521	(89)	ACG→A-G	FS	12542	(81)	ACT→ATT	T→I	13080	(81)	TTC→TTT	F (0.55→0.45)
13233,4	(42)	TCC/TTT→TC G/CTT	S,F→S,L	14416	(6)	CCT→CTT	P→L	14665	(38)	GCT→GTT	A→V
14813	(92,4)	TAC→TAT	Y (0.57→0.43)	15201	(91)	GCT→-CT	FS	15332	(39)	AAC→AAT	N (0.54→0.46)
15594	(91)	GCC→-CC	FS	15605	(48)	TAT→TAC	Y→Y	15605	(48)	ATG→ACG	M→T
15615	(64)	TTA→CTA	L (0.07→0.07)	15818	(91)	AAC→AA-	FS	15818	(91)	AAC→AA-	FS
16475	(77)	CCA→CCG	p (0.27→0.11)	16885	(74)	ACA→ATA	T→I	17008	(57)	ACA→ATA	T→I
17255	(4)	CGT→CGC	R (0.08→0.19)	17381	(38, 47, 63, 14)	GCC→GCT	A (0.40→ 0.26)	17384	(42)	ACA→ACG	T (0.28→0.12)
17418	(3)	CGT→TGT	R→C	17431	(46)	TAT→TGT	Y→C	17755	(72,73,76-78,80)	CCT→CTT	P→L
17866	72,73,76-78,80)	TAT→TGT	Y→C	18068	(72,73,76-78,80,84,85)	CTC→CTT	L(0.20→0.13)	18611	(82)	CAT →CAC	H(0.41→0.59)
18822	(6)	CTG→TTG	L(0.41→0.13)	18983	(82)	GTT→GTA	V (0.18→ 0.11)	19073	(5)	CCT→CCC	P(0.28→0.33)
19073	(5)	CCT→CCC	P(0.28→0.33)	19183	(82)	GAT→GCT	D→A	19618	(14)	ACA→ATA	T→I
20289	(76)	TTC→CTC	F→L	20306-8	(47)	-AAA/TTA→A A	L Deleted	20944	(48)	ACG→ATG	T→M
20996	(89)	ACT→AC-	FS	21145	(9)	AAG-AGG	K→R	21155	(78)	CTT→CTC	L (0.13→ 0.20)
21324	(11)	GAT→AAT	D→N	21392,5	(77)	T-T→TTTCT	F added	21397	(77,78)	TCT→TTT	S→F
S protein											
21655	(67)	TAC→AAC	Y→N	21718	(50)	CAT→TAT	H→Y	22002,4	(38)	GTT/TAT→T-T	Y deleted
22044	(25)	TTC→TTA	F→L	22115	(87)	GGA→GTA	G→V	22235	(48)	TCG→TGG	S→W
22314	(5)	AGT→AGG	S→R	22443	(78)	GAC→GAT	D (0.54→0.46)	22796	(38)	AGA→ATA	R→I
23196	(77)	TTC→TTT	F (0.55→0.45)	23414	(6)	GAT→GGT	D→G	23963	(42)	TTT→TGT	F→C
24045	(62)	AAC→AA-	FS	24045	(43, 46, 71, 66)	AAC→AAT	N (0.54→0.46)	24336	(11,41)	AAA→AAG	K (0.42→ 0.58)
24362	(74)	GCT→GTT	A→V								
Orf3a											
25598	(89)	CTC→CT-	FS	25786	(48)	TGG→TTG	W→L	25821	(60)	CTT→GTT	L→V
26155	(92, 42, 48, 4, 90, 57, 5)	GGT→GTT	G→V								
E protein											
26337	(19, 36)	CTA→TTA	L (0.07→0.07)	26365	(48)	CTT→CAT	L→H				
M protein											
26740	(66, 71)	GCT→GCC	A (0.26→0.40)	26740	(62)	GCT→GC-	FS				
Orf8											
27936	(82)	ACA→ATA	T→I	28088	(66, 71)	GTG→CTG	V→L	28088	(62)	GTG→-TG	FS
28155	(61,64-84)	TTA→TCA	L→S	28155	(62)	TTA→T-A	FS	28264	(91)	TTC→TT-	FS
28378	(89)	ACG→A→G	FS	28378	(89)	CGC→-GC	FS	28389	(24)	CGC→CTC	R→L
28389	(24,91)	GCG→GCT	A→A	28420	(23)	CCC→TCC	P→S	28803	(71)	GCA→GCT	A (0.23→ 0.26)
28865	(62)	TCA→T-A	FS	28865	(51, 46)	TCA→TTA	S→L	28889	(83)	AGT→AAT	S→N
28927	(89)	GGT→AGT	G→S	29106	(65, 64, 82,79)	TTC→TTT	F (0.55→ 0.45)	29241	(21)	CGC→CGT	R(0.19→0.08)
29312	(70)	GAT→GTT	D→V	29314	(60, 39)	CCA→TCA	P→S	29538	(14,63)	CAG→CAA	Q(0.75→0.25)
Seq. number	Ref.	Mutation	Result	Seq. number	Ref.	Mutation	Result	Seq. number	Ref.	Mutation	Result
Seq. number	Ref.	Mutation	Result	Seq. number	Ref.	Mutation	Result	Seq. number	Ref.	Mutation	Result
Orf10											
29646	(86, 21, 32)	TAC→TAT	Y (0.57→ 0.43)								
3'UTR											
29747	(23,87)	G→T	-	29753	(83)	G→A	-	29761,70	(5)	ACGATCGAGTG T→A-T	-
29762	(87)	G→C	-								

Table 4. Phylogenetic tree analyzing of 93 whole genome sequences of SARS-CoV-2 shows the new subdivisions of IA and L types of this virus, the leftover sequences were belong to L1 sub division.



References

1. Poland GA. Another coronavirus, another epidemic, another warning. Vaccine. 2020; 38: v–vi.
2. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020;26:450-2.
3. Ren L-L, Wang Y-M, Wu Z-Q, Xiang Z-C, Guo L, Xu T, et al. Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. Chin Med J. 2020;133:1015-1024.
4. Jiang S, Du L, Shi Z. An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies. Emerg Microbes Infect. 2020;9:275-7.
5. Wu Y, Ho W, Huang Y, Jin D-Y, Li S, Liu S-L, et al. SARS-CoV-2 is an appropriate name for the new coronavirus. The Lancet. 2020;395:949-50.
6. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet. 2020; 395:565-74.
7. World Health Organization (WHO). Coronavirus disease 2019 (COVID-19): situation report, 72. 2020.
8. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev. 2020;7:1012–1023.
9. Gu H, Chu DK, Peiris JSM, Poon LL. Multivariate Analyses of Codon Usage in 2019 Novel Coronavirus on the Genomic Landscape of Betacoronavirus. bioRxiv. 2020.
10. Zhang L, Yang J-R, Zhang Z, Lin Z. Genomic variations of SARS-CoV-2 suggest multiple outbreak sources of transmission. medRxiv. 2020.
11. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah N, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res. 2020;176:104742.
12. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. Curr Biol. 2020;30:1578.
13. Yang P, Wang X. COVID-19: a new challenge for human beings. Cell Mol Immunol. 2020;17:555-557.
14. Li X, Zai J, Zhao Q, Nie Q, Li Y, Foley BT, et al. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J Med Virol. 2020; 92:602-611.
15. Lam TT-Y, Shum MH-H, Zhu H-C, Tong Y-G, Ni X-B, Liao Y-S, et al. Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. BioRxiv. 2020.
16. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou J-J, et al. Isolation and characterization of 2019-nCoV-like coronavirus from Malayan pangolins. BioRxiv. 2020.
17. Kannan S, Ali PSS, Sheeza A, Hemalatha K. COVID-19 (Novel Coronavirus 2019)–recent trends. Eur Rev Med Pharmacol Sci. 2020;24:2006-11.
18. Patiño-Galindo JÁ, Filip I, AlQuraishi M, Rabadan R. Recombination and convergent evolution led to the emergence of 2019 Wuhan coronavirus. bioRxiv. 2020.
19. Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). Clin Exp Pediatr. 2020;63:119.
20. Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, et al. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. Viruses. 2020; 12:244.
21. Silvestro D, Michalak I. RaxmlGUI: a graphical front-end for RAxML. Org Divers Evol. 2012; 12:335-7.
22. Cao Y, Li L, Feng Z, Wan S, Huang P, Sun X, et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discov. 2020;6:1-4.
23. Liu Z, Xiao X, Wei X, Li J, Yang J, Tan H, et al. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. J Med Virol. 2020;92: 595-601.
24. Koyama T, Platt D, Parida L. Variant analysis of COVID-19 genomes. Bull World Health Organ. E-pub: 24 February 2020.