Research Article

Corona virus (CoVid19) genome: genomic and biochemical analysis revealed its possible synthetic origin

Abstract

The Severe acute respiratory syndrome (SARS) corona virus 2 SARS-CoV-2 mediated epidemic is a global pandemic. It has evolved as a curse to the human civilization and at the present situation, where most of the cities in the world are on lockdown. The first genome sequence data of SARS-CoV-2 (CoVid19) and their reports that followed concluded that it was a member of the genus Betacoronavirus and has a bat reservoir. To understand its origin and evolution, we conducted a deep comparative study by comparing the genomes of bat SARS CoV and other SARS CoVs (including human SARS CoV of German isolate). Results revealed that CoVid19 genomes from isolates of China, India, Italy, Nepal, and the United States of America has sequence similarity of 79-80% only with the bat SARS CoV and it has sequence similarity of approximately 60% with the human SARS CoV of German isolate. Whereas, the sequence similarity within the CoVid19 genomes of these countries was 99-100%. If the SARS CoV infection happened to human through the SARS CoV of bat origin, it should have sequence similarity of more than 99% which was absent in this case. Phylogenetic analysis revealed, bat SARS CoV did not fall with the group of SARS CoV of China, India, Italy, Nepal, and USA isolates. The genome analysis revealed the presence of multiple microsatellite repeats sequences. Proteome analysis revealed, the melting temperature (Tm) of surface glycoprotein was less than 55°C, suggesting the steam treatment can be an ideal preventative measure to destabilize the CoVid19, and thus it's spreading.

Keywords: SARS, corona virus, SARS-CoV-2, CoVid19, MERS, epidemic, pandemic

Tapan Kumar Mohanta,[1] Yugal Kishore Mohanta[2]
[1]Natural and Medical Sciences Research Center, University of Nizwa, Oman
[2]Department of Botany, North Orissa University, India

Correspondence: Tapan Kumar Mohanta, Natural and Medical Sciences Research Center, University of Nizwa, Nizwa, 616, Oman, Email nosoc.tapan@gmail.com, taan.mohanta@unizwa.edu.om

Abbreviations: SARS, severe acute respiratory syndrome; Tm, temperature; CoV, corona virus

Introduction

Severe acute respiratory syndrome (SARS) corona virus 2 belonged to the family Coronavriridae of the order Nidovirales. It contains a positive-sense single stranded genome. The genome encodes overlapping polyproteins ORF1ab, surface glycoprotein, ORF3a, envelope protein, membrane glycoprotein, ORF6, ORF7a, ORF8, nucleocapsid phosphoprotein, and ORF10. The ORF1ab get processed into the viral polymerase. The Corona virus (CoV) causes disease in variety of the wild and domestic animals including humans. The α- and β-CoVs usually infect the mammals and γ-and δ-CoVs infect birds.[1] The β-CoVs including middle east respiratory syndrome (MERS)- CoV and sever acute respiratory syndrome (SARS)-CoVs caused global pandemic since 2002-2003. The SARS-CoV was originated from China and created a global pandemic by infecting more than 8000 individual with mortality rate of 10%.[2] Since then, the disease related to CoVs became a dangerous threat to the human civilization. Recently, SARS-CoV-2 (known as CoVid19) had outbreak from the Wuhan city of China and got spread all over the world by infecting more than 238 million people with more than 817000 deaths so far (25th August 2020). This resulted a severe global pandemic and shaken the economy of the world. At the present situation, the whole world is locked down to stop the human-to-human spread of SARS-CoV-2 and we do not know when we will overcome this situation. The lack of medicine or vaccine at the present situation made it uncontrolled. Although several research laboratories of the world are actively working in different strategies to control the SARS-

CoV-2, there are several misconception and misrepresentation about the genomics, mutation, and evolution of the SARS-CoV-2 genome in the public domain. Few of the general misconception regarding the SARS-CoV-2 are as follows; SARS-CoV-2 has evolved from bat or pangolin corona virus,[3] two CoV genome merged to become SARS-CoV-2, SARS-CoV-2 was synthesized in the laboratory to use as a bio-weapon, the effectiveness of the SARS-CoV-2 in cold weather country is high, garlic and hot water reduces the effectiveness of SARS-CoV-2 and others. To address these highly important aspects, we have conducted genomic, proteomic, and evolutionary study to understand the mutation and biochemical features of the SARS-CoV-2 genomes and proteomes.

Results

SARS-CoV-2 genome do not have significant similarity with bat or pangolin SARS corona virus

The genome size of SARS-CoV-2 genome varies from 27317 to 29903 nucleotides with GC content of 38 to 38.3% (Table 1). The SARS-CoV-2 genome encodes for 10 to 12 complementary DNA sequences (CDS). The isolate of China Wuhan-Hu-1 (accession NC_045512.2) contained 12 CDS whereas other SARS-CoV-2 genomes encode 10 CDS (Table 1). The human SARS corona virus reported in 2003 in Germany also encoded 12 CDS (Table 1). To understand the genomic and evolutionary aspects of SARS-CoV-2 genome, we downloaded 24 whole genome sequences of SARS-CoV-2 originated from different countries of the world. This includes 12 SARS-CoV-2 genomes from China, seven from the United States of America (USA), and one each from Canada, Germany, India, Italy,

Nepal, and the United Kingdom. The human corona virus genome of Germany (accession number NC_004718.3) was also considered as a reference for the comparative study as it was reported long back in 2003.[4] We made a comparative sequence similarity study of recent SARS-CoV-2 genomes with isolates of human corona virus of German origin. We found SARS-CoV-2 genome had sequence similarity of 60.13% (Nepal CoVid19) to 60.24% (China SARS-CoV-2 Wuhan-Hu-1) with the German origin human corona virus (Table 2). Comparative similarity study of SARS-CoV-2 genomes with bat SARS KHU3-1 showed similarity level ranged from 79.44% (CoVid19 USA) to 79.78 % (CoVid19 China Wuhan-Hu-1) (Table 2). Comparative similarity study of human SARS-CoV-2 with bat SARS WIV1 showed 60.01% (SARS CoV Germany) to 80.10% (SARS-CoV-2 China Wuhan-Hu-1 and CoVid19 USA). The MERS CoV showed sequence similarity of 54.59% with SARS-CoV-2 India to 61.25% with SARS CoV-2 of China Wuhan-Hu-1. The SARS-CoV-2 of Wuhan-Hu-1 was originated recently from Wuhan and found in CoVid19 patients (Table 2). Therefore, we made a comparative study of SARS-CoV-2 Wuhan-Hu-1 isolates with SARS-CoV-2 isolates of other countries. We found SARS-CoV-2 Wuahn-Hu-1 had 100% sequence similarity with the SARS-CoV-2 isolates of Nepal followed by similarity level of 99.99% (Italy and the USA), 99.98% (India), and 60.24% (Germany) (Table 2). Till date (4th April 2020) there was presence of 12 CoVid19 genome sequences from the Chinese origin. Therefore, we made a comparative study by aligning all the full-length genome sequences of all the 12 Chinese SARS-CoV-2 genomes. The Chinese CoVid19 Wuhan-Hu-1 has 5'-untranslated region (UTR) from 1st to 265 nucleotides and 3'-UTR from 29675 to 29903 nucleotides. Alignment showed, there were slight differences in the 5' and 3'-UTR and no mutation/substitution was found in the open reading frame (ORF) in the SARS-CoV-2 genomes of Chinese isolates (Supplementary Figure 1). Similarly, there was seven SARS-CoV-2 isolates of the United States of America (USA) origin. We aligned all the seven CoVid19 genomes of isolates of the USA to find the possible mutation or substitution in them. Resulted showed, all the seven genomes had 100% sequence similarity and no mutation or substitution was found within them (Supplementary Figure 2). All the 5' and 3' UTRs were also found to be conserved (Supplementary Figure 2). Later, we aligned recently reported SARS-CoV-2 sequences of China Wuhan-Hu-1, India, Italy, Nepal, and USA. The SARS-CoV-2 of Indian isolate has substituted/mutated G instead of A at position 1671, Italian isolate had substituted T instead of A at position 2269, Indian CoVid19 had substituted T in the place of Cat position 6481, India and USA had substituted T instead of Cat position 8762 and 8782, respectively, Italy has unknown nucleotide N instead of G at position 11083, India had T instead of C at position 16857, Nepal has substituted T instead of C at position 24019, India has T instead of C at position 24331, and Italy has substituted T instead of G at position 26144 (Supplementary Figure 3).

Table 1 Genomic details of different isolates of SARS corona virus 2 from different countries of the world

SARS-CoV-2 strain	Accession number	Genome size (Mb)	GC content (%)	Number of proteins
China (Wuhan-Hu-1)	MN908947.3	0.029903	38	10
China (Wuhan-Hu-1)	NC_045512.2	0.029903	38	12
India	MT050493.1	0.029851	38	10
Germany	NC_002645.1	0.027317	38.3	12
Italy	MT066156.1	0.029867	38	10
Nepal	MT072688.1	0.029811	38	10
United States of America	MN985325.1	0.029882	38	10

Table 2 Comparative genomic analysis of SARS-CoV-2 isolates from different countries with SARS corona virus of source organism

Corona virus isolate from Country	Accession number	Similarity with human CoVid/ German (%)	Similarity with Bat SARS HKU3-1 (%)	Similarity with Bat SARS WIV1 (%)	MERS Corona Virus (%)	Similarity with SARS Cov2 Wuhan-Hu-1 (%)
China (Wuhan-Hu-1)	MN908947.3	60.24	79.78	80.1	61.25	*****
India	MT050493.1	60.15	79.74	80.07	54.59	99.98
Germany	NC_002645.1	****	59.77	60.01	60.24	60.24
Italy	MT066156.1	60.17	79.76	80.08	61.23	99.99
Nepal	MT072688.1	60.13	79.74	79.67	54.65	100
United States of America	MN985325.1	60.19	79.44	80.1	61.15	99.99

SARS CoVid19 genome is closer to the Bat SARS corona virus genome

To understand the evolutionary linkage of human SARS-CoV-2 with bat SARS CoV and other SARS CoV, we constructed a phylogenetic tree by considering the whole genome sequences of the SARS CoVs. In the study, there were five SARS CoV2 (CoVid19) isolates from different countries whose genome was reported recently. In addition, there was genome sequences of bat CoV, beta SARS CoV of Canada, MERS CoV, United Kingdom beta CoV, bovine CoV, and human CoV 229E (German isolate) as well. The bat SARS CoV HKU3-1 and bat SARS CoV WIV-1 were found close to the human SARS-CoV-2, but fall in a separate group (Figure 1). The bat SARS CoV genome did not group with the SARS CoV2 (CoVid19). However, none of the other CoVs were found closer to the human SARS-CoV-2. The time tree analysis of SARS CoV2 genomes revealed their origin from 0.00 million years ago suggesting their recent origin (Figure 2). The recombination events of the SARS-CoV-2 with other SARS CoV genomes showed no recombination event within themselves or between other SARS CoVs (Figure 3). To understand the nucleotide substitution, a maximum composite likelihood estimate of the pattern of nucleotide substitution was conducted. It showed higher rate of transition compared to the transversion (Table 3). The substitution of T to C nucleotide was 58.98 and the substitution of C to T nucleotide was 34.06 (Table 3). The substitution of purines A to G nucleotide was one and substitution of G to A nucleotide was 0.72. The substitution of A to C/T or G to C/T nucleotide and vice versa was less than one (Table 3). The transition rate of SARS-CoV-2 genome of isolates from China Wuhan-Hu-1, India, Italy, Nepal and USA from C to T nucleotide was 26.82 whereas the transition from T to C nucleotide was 46.86 (Table 3). However, the transversion rate was found below 3 (Table 3).
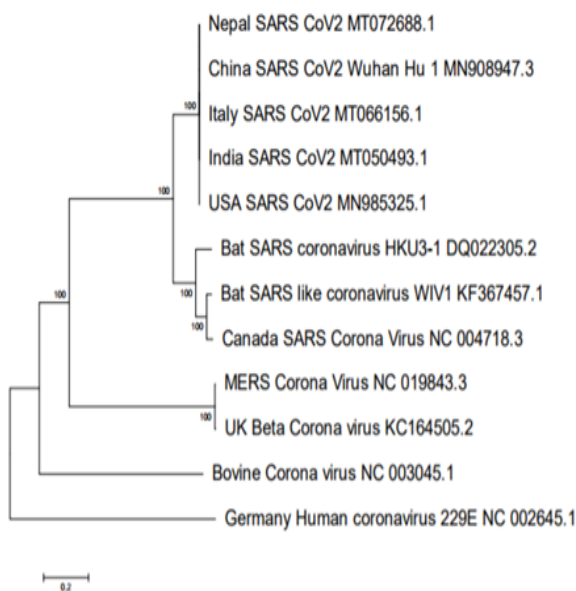

Figure 1 Phylogenetic tree of SARS-CoV-2 genomes of China, India, Italy, Nepal, and the United States of America with bat SARS CoV, human SARS CoV (229E German isolates), MERS CoV and others. The SARS-CoV-2 genomes grouped together whereas other SARS CoV genome fall in the different group. The possible donor bat SARS CoV do not group with CoVid19 genomes, suggesting the origin of CoVid19 from other sources. The phylogenetic was constructed using MEGA6.

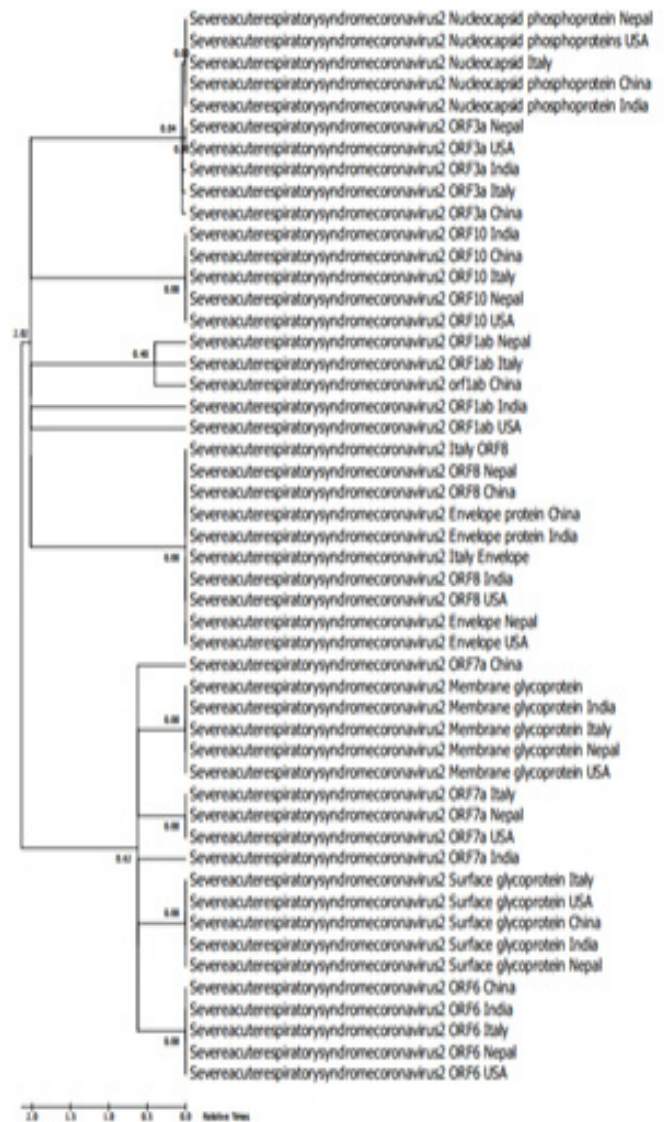**Figure 2** Time tree analysis of SARS-CoV-2 genomes with other SARS CoVs. Analysis showed evolution of SARS-CoV-2 genome to 0.00 million years, suggesting their recent origin. The time tree analysis was conducted in MEGA6.
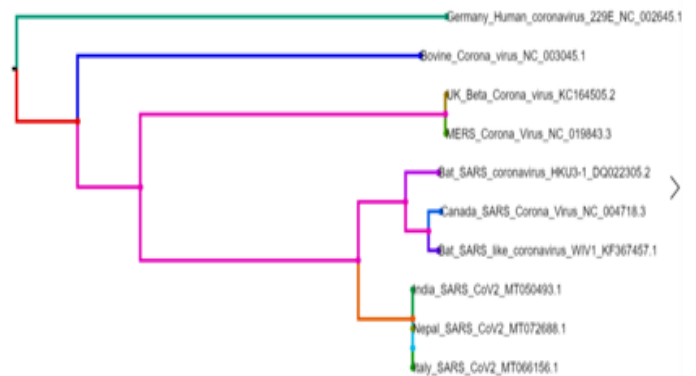

Figure 3 Recombination event of SARS CoVs with SARS-CoV-2 genomes. Result showed, SARS CoVs did not undergone recombination within them suggesting the presence of genome sequences from other/synthetic donor.

Table 3 Maximum composite likelihood estimate of the pattern of nucleotide substitution of SARS CoV genomes

	A	T	C	G
A	-	0.85	0.49	0.72
T	0.75	-	34.06	0.54
C	0.75	58.98	-	0.54
G	I	0.85	0.49	-

Substitution of SARS-CoV-2 Isolates of China, India, Italy, Nepal, and USA isolates

	A	T	C	G
A	-	2.77	1.58	3.6
T	2.57	-	26.82	1.69
C	2.57	46.86	-	1.69
G	5.48	2.77	1.58	-

SARS-CoV-2 genome contain microsatellite repeats

Microsatellites are the repetitive DNA motifs of length ranged from one to six or more nucleotides. Analysis revealed the presence of at least 34 unique microsatellites repeat sequences in SARS-CoV-2 genome (Supplementary Table 1). The microsatellite repeats sequences TGTGTG and ACACAC were found 12 times, GTGTGT nine times, ATATAT, and CACACA eight times (Supplementary

Table 1). The microsatellites sequences were mapped with the CDS of CoVid19 genome and it was found in the ORF1ab, surface glycoprotein, envelope protein, ORF3a, nucleocapsid phosphoprotein. The microsatellite repeats sequence GTGTGTGTGT found at the position 20486 did not mapped to the CDS, suggesting its occurrence in the non-coding region. The ORF6, ORF7a and ORF8 did not have any microsatellite repeats. The microsatellites present in the coding region might cause phenotypic change and disease.

Few CoVid19 proteins undergone amino acid substitution/mutation

Multiple sequence alignment revealed, a few SARS-CoV-2 proteins have undergone substitution/mutation. In the ORF1ab of Indian isolate amino acid P (proline) was substituted to L (leucine) at the position 2079 and amino acid T (threonine) was substituted for I (isoleucine) at position 5538 of the protein sequence (Table 4) (Supplementary Figure 4). However, in ORF1ab of isolate of Italy, amino acid L was substituted for X (unknown) at position 3606 (Table 4) (Supplementary Figure 4). In the surface glycoprotein of Indian isolate, amino acid A (alanine) at the position 929 was substituted for V (valine) (Supplementary Figure 5). In ORF3a, amino acid G (glycine) was substituted at position 251 for V in Italian isolate (Table 4) (Supplementary Figure 6). In ORF8, amino acid L was substituted for S (serine) at position 84 in Indian and USA isolates (Table 4) (Supplementary Figure 7). No mutation or substitution was observed for envelope protein, membrane glycoprotein, nucleocapsid phosphoprotein, ORF6, ORF7a, and ORF10.

Supplementary Table 1 Microsatellite repeats of SARS-CoV-2 genome

Repeats in CoVid19	Total No. of Repeats	Position	Mapped in ORF
TGTGTG	12	84, 1489, 2327, 4438, 10844, 11546, 14827, 15442, 15728, 16483, 20486, 26359	ORF1ab, Envelope protein,
ACACAC	12	298, 4571, 6188, 8954, 9116, 10999, 12917, 13162, 13661, 16213, 18111, 18553	ORF1ab, Surface glycoprotein
TTCTTCTTC	2	626, 22320	ORF1ab, Surface glycoprotein,
AAAAAA	7	1813, 11990, 29870	ORF1ab
GTGTGT	9	2421, 5515, 17508, 19055, 20486, 21603, 24654, 27458, 29687	ORF1ab, Surface glycoprotein
GAAGAAGAA	2	3055, 3073	ORF1ab
AAGAAGAAG	2	3188, 29389	ORF1ab
GATGATGAT	I	3205	ORF1ab
ATATAT	8	4116, 7254, 11727, 13777, 13948, 19903, 22168, 29593	ORF1ab, Surface glycoprotein, ORF10
TATATA	6	4237, 16510, 22664, 25186, 26660, 29563	ORF1ab, Surface glycoprotein, ORF10
TCTCTC	5	4666, 7813, 18566, 22073, 25147	ORF1ab, Surface glycoprotein
CTTCTTCTT	2	4736, 14756	ORF1ab
AGAGAG	5	4850, 6121, 14270, 14484, 22954	ORF1ab, Surface glycoprotein
GAGAGA	3	4950, 7674, 22954	ORF1ab, Surface glycoprotein
CACACA	8	5170, 6538, 13162, 19151, 19317, 24858, 26130, 29545	ORF1ab, Surface glycoprotein, ORF3a,
TCTCTCTCTC	I	7813	ORF1ab

Table Continued...

Repeats in CoVid19	Total No. of Repeats	Position	Mapped in ORF
TTTTTT	4	9627, 11074, 19983, 21101	ORF1ab
TTTTTTTT	1	11074	ORF1ab
ATGATGATG	1	11366	ORF1ab
ATCATCATC	1	11910	ORF1ab
ACACACAC	1	13162	ORF1ab
TGATGATGA	1	13895	ORF1ab
CTCTCT	4	7813, 15711, 17122, 22445,	ORF1ab, Surface glycoprotein
GTGTGTGTGT	1	20486	NA
GAGAGAGA	1	22954	Surface glycoprotein
AGTAGTAGT	1	23088	Surface glycoprotein
TGTTGTTGT	1	25642	ORF3a
AATAATAAT	1	25757	OR3a
CGACGACGA	1	26191	ORF3a
GTGGTGGTG	1	28556	Nucleocapsid phosphoprotein
TGCTGCTGC	1	28934	Nucleocapsid phosphoprotein
CAACAACAACAA	1	28987	Nucleocapsid phosphoprotein
CTGCTGCTG	1	29021	Nucleocapsid phosphoprotein
AAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAA	1	29870	NA

Table 4 Substitution of SARS corona virus SARS-CoV-2 proteins of isolates

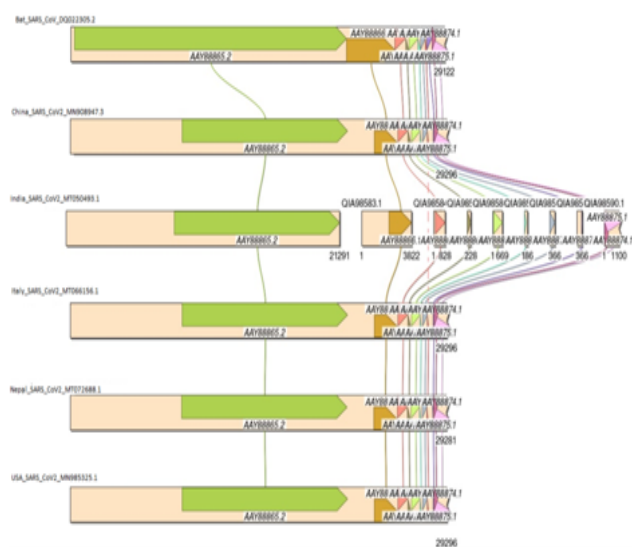| Name of the protein	Substitution (position in the sequence)	Substituted amino acid	Isolate of the Country
Envelope protein	NA	NA	NA
Membrane glycoprotein	NA	NA	NA
Nucleocapsid phosphoprotein	NA	NA	NA
ORF1ab	2079	P > L	India
	3606	L > X	Italy
	5538	T > I	India
Surface glycoprotein	929	A > V	India
ORF3a	251	G > V	Italy
ORF6	NA	NA	NA
ORF7a	NA	NA	NA
ORF8	84	L > S	India
	84	L > S	USA
ORF10	NA	NA	NA

Figure 4 Principal component analysis of amino acid composition of SARS-CoV-2 genomes of China, India, Italy, Nepal, and USA isolates. Amino acid Leu has highest abundance and fall apart from other amino acids.

The melting temperature (Tm) of membrane glycoprotein is less than 55°C

We studied the Tm of all the ten proteins found in the genome of SARS-CoV-2 (CoVid19). Analysis revealed, the Tm of the membrane glycoprotein was less than 55°C. The Tm of ORF1ab, surface glycoprotein, ORF3a, envelope protein, and nucleocapsid phosphoprotein was found 55-65°C (Supplementary Figure 8). However, the Tm of ORF6, ORF7a, and ORF10 was found greater than 65°C (Supplementary Table 2). The half-life period of all the proteins were found above 30 hours for reticulocytes/in vitro and more than 20 hours for in vivo (Supplementary Table 3). All the proteins were also found to be stable and the stability of the nucleocapsid phosphoprotein was highest (instability index 55.09). The stability of nucleocapsid phosphoprotein was followed by ORF7a, ORF8, membrane glycoprotein, envelope protein, ORF1ab, surface glycoprotein, ORF3a, ORF6, and ORF10 (Supplementary Table 3).

Amino acid composition of leu was highest and Trp was lowest in CoVid19 proteome

To understand the amino acid composition, we analysed all the full-length protein sequences of the SARS-CoV-2 proteomes. We found Leu (9.489%) was the highest and Trp (1.118%) was the lowest abundant amino acid in the SARS-CoV-2 proteome (Supplementary Table 4). The highest abundance of Leu amino acids in CoVid19 proteome was followed by Val (8.084%), Thr (7.428%), and Ser (6.785%) (Supplementary Table 4). Principal component analysis of amino acid composition revealed the grouping of Asn, Tyr, Thr, Phe, and Ser; Pro, Gly, Arg, and Cys; and Trp, His, Gln, Asp, Lys, and Glu (Figure 4). The ORF1ab encodes for highest number (7096) of amino acids whereas ORF10 encodes lowest number of amino acids (38) (Supplementary Figure 9).

Supplementary Table 2 Predicted melting temperature (Tm) of SARS-CoV-2 proteins

Protein ID	Protein Name	Tm (oC)	Tm Index
China			
YP_009724389.1	ORF1ab Polyprotein	55-65	0.563
YP_009725295.1	ORF1a Polyprotein	55-65	0.461
YP_009724390.1	Surface Glycoprotein	55-65	0.464
YP_009724391.1	ORF3a	55-65	0.224
YP_009724392.1	Envelope	55-65	0.52
YP_009724393.1	Membrane Glycoprotein	< 55	-0.34
YP_009724394.1	ORF6	> 65	1.055
YP_009724395.1	ORF7a	> 65	2.968
YP_009725318.1	ORF7b	< 55	-0.82
YP_009724396.1	ORF8	> 65	1.465
YP_009724397.2	Nucleocapsid Phosphoprotein	55-65	0.318
YP_009725255.1	ORF10 Protein	> 65	1.637
India			
QIA98582.1	ORF1ab Polyprotein	55-65	0.567
QIA98583.1	Surface Glycoprotein	55-65	0.461
QIA98584.1	ORF3a	55-65	0.224
QIA98585.1	Envelope Protein	55-65	0.52
QIA98586.1	Membrane Glycoprotein	< 55	-0.34
QIA98587.1	ORF6	> 65	1.055
QIA98588.1	ORF7a	> 65	1.773

Table Continued...

Protein ID	Protein Name	Tm (oC)	Tm Index
India			
QIA98589.1	ORF8	> 65	1.465
QIA98590.1	Nucleocapsid Phosphoprotein	55-65	0.318
QIA98591.1	ORF10	> 65	1.637
Italy			
QIA98553.1	ORF1ab polyprotein		
QIA98554.1	surface glycoprotein	55-65	0.464
QIA98555.1	ORF3a	55-65	0.328
QIA98556.1	envelope protein	55-65	0.52
QIA98557.1	membrane glycoprotein	< 55	-0.34
QIA98558.1	ORF6	> 65	1.055
QIA98559.1	ORF7a	> 65	1.773
QIA98560.1	ORF8	> 65	1.465
QIA98561.1	nucleocapsid phosphoprotein	55-65	0.318
QIA98562.1	ORF10	> 65	1.637
Nepal			
QIB84672.1	ORF1ab polyprotein	55-65	0.563
QIB84673.1	surface glycoprotein	55-65	0.464
QIB84674.1	ORF3a	55-65	0.224
QIB84675.1	Envelope protein	55-65	0.52
QIB84676.1	membrane glycoprotein	< 55	-0.34
QIB84677.1	ORF6	> 65	1.055
QIB84678.1	ORF7a	> 65	1.773
QIB84679.1	ORF8	> 65	1.465
QIB84680.1	nucleocapsid phosphoprotein	55-65	0.318
QIB84681.1	ORF10	> 65	1.637
United States of America			
QHO60603.1	ORF1ab polyprotein	55-65	0.563
QHO60594.1	surface glycoprotein	55-65	0.464
QHO60595.1	ORF3a	55-65	0.224
QHO60596.1	envelope protein	55-65	0.52
QHO60597.1	membrane glycoprotein	< 55	-0.34
QHO60598.1	ORF6	> 65	1.055
QHO60599.1	ORF7a	> 65	1.773
QHO60600.1	ORF8	> 65	1.465
QHO60601.1	nucleocapsid phosphoprotein	55-65	0.318
QHO60602.1	ORF10	> 65	1.637

Citation: Mohanta TK, Mohanta YK. Corona virus (CoVid19) genome: genomic and biochemical analysis revealed its possible synthetic origin. *J Appl Biotechnol Bioeng.* 2020;7(5):200‒213. DOI: 10.15406/jabb.2020.07.00235

Supplementary Table 3 Half-life period and instability index of SARS-CoV-2 proteins

Proteins	Isolate Country	Molecular formula	Half-life in reticulocytes/vitro (Hrs)	Half-life in vivo (Hrs)	Instability Index (II)
Envelope protein	China	C390H625N91O103S4	30	> 20	38.68/Stable
Envelope protein	India	C390H625N91O103S4	30	> 20	38.68/Stable
Envelope protein	Italy	C390H625N91O103S4	30	> 20	38.68/Stable
Envelope protein	Nepal	C390H625N91O103S4	30	> 20	38.68/Stable
Envelope protein	USA	C390H625N91O103S4	30	> 20	38.68/Stable
Membrane glycoprotein	China	C1165H1823N303O301S8	30	> 20	39.14/Stable
Membrane glycoprotein	India	C1165H1823N303O301S8	30	> 20	39.14/Stable
Membrane glycoprotein	Italy	C1165H1823N303O301S8	30	> 20	39.14/Stable
Membrane glycoprotein	Nepal	C1165H1823N303O301S8	30	> 20	39.14/Stable
Membrane glycoprotein	USA	C1165H1823N303O301S8	30	> 20	39.14/Stable
Nucleocapsid phosphoprotein	China	C1971H3137N607O629S7	30	> 20	55.09/Stable
Nucleocapsid phosphoprotein	India	C1971H3137N607O629S7	30	> 20	55.09/Stable
Nucleocapsid phosphoprotein	Italy	C1971H3137N607O629S7	30	> 20	55.09/Stable
Nucleocapsid phosphoprotein	Nepal	C1971H3137N607O629S7	30	> 20	55.09/Stable
Nucleocapsid phosphoprotein	USA	C1971H3137N607O629S7	30	> 20	55.09/Stable
ORF1ab	China	C35644H55333N9253O10496S394	30	> 20	33.31/Stable
ORF1ab	India	C35646H55339N9253O10495S394	30	> 20	33.25/Stable
ORF1ab	Italy	C35638H55322N9252O10495S394	30	> 20	33.36/Stable
ORF1ab	Nepal	C35644H55333N9253O10496S394	30	> 20	33.31/Stable
ORF1ab	USA	C35644H55333N9253O10496S394	30	> 20	33.31/Stable
ORF3a	China	C1440H2189N343O404S11	30	> 20	32.96/Stable
ORF3a	India	C1440H2189N343O404S11	30	> 20	32.96/Stable
ORF3a	Italy	C1443H2195N343O404S11	30	> 20	32.96/Stable
ORF3a	Nepal	C1440H2189N343O404S11	30	> 20	32.96/Stable
ORF3a	USA	C1440H2189N343O404S11	30	> 20	32.96/Stable
ORF6	China	C334H532N78O96S3	30	> 20	31.16/Stable
ORF6	India	C334H532N78O96S3	30	> 20	31.16/Stable
ORF6	Italy	C334H532N78O96S3	30	> 20	31.16/Stable
ORF6	Nepal	C334H532N78O96S3	30	> 20	31.16/Stable
ORF6	USA	C334H532N78O96S3	30	> 20	31.16/Stable
ORF7a	China	C633H988N156O171S7	30	> 20	48.66/Stable
ORF7a	India	C633H988N156O171S7	30	> 20	48.66/Stable
ORF7a	Italy	C633H988N156O171S7	30	> 20	48.66/Stable
ORF7a	Nepal	C633H988N156O171S7	30	> 20	48.66/Stable
ORF7a	USA	C633H988N156O171S7	30	> 20	48.66/Stable
ORF8	China	C633H961N155O177S8	30	> 20	45.79/Stable

Table Continued...

Proteins	Isolate Country	Molecular formula	Half-life in reticulocytes/vitro (Hrs)	Half-life in vivo (Hrs)	Instability Index (II)
ORF8	India	C630H955N155O178S8	30	> 20	46.24/Stable
ORF8	Italy	C633H961N155O177S8	30	> 20	45.79/Stable
ORF8	Nepal	C633H961N155O177S8	30	> 20	45.79/Stable
ORF8	USA	C630H955N155O178S8	30	> 20	46.24/Stable
ORF10	China	C206H312N50O54S3	30	> 20	16.06/Stable
ORF10	India	C206H312N50O54S3	30	> 20	16.06/Stable
ORF10	Italy	C206H312N50O54S3	30	> 20	16.06/Stable
ORF10	Nepal	C206H312N50O54S3	30	> 20	16.06/Stable
ORF10	USA	C206H312N50O54S3	30	> 20	16.06/Stable
Surface glycoprotein	China	C6336H9770N1656O1894S54	30	> 20	33.01/Stable
Surface glycoprotein	India	C6338H9774N1656O1894S54	30	> 20	33.01/Stable
Surface glycoprotein	Italy	C6336H9770N1656O1894S54	30	> 20	33.01/Stable
Surface glycoprotein	Nepal	C6336H9770N1656O1894S54	30	> 20	33.01/Stable
Surface glycoprotein	USA	C6336H9770N1656O1894S54	30	> 20	33.01/Stable

Supplementary Table 4 Amino acid composition of SARS-CoV-2 from different countries of the word

Amino	SARS-CoV-2 Sequences from Different Countries					Average (%)
Acids	China	India	Nepal	Italy	USA	
Ala	656	655	637	656	656	6.772
Cys	294	294	290	294	294	3.045
Asp	509	509	503	509	509	5.219
Glu	439	439	432	439	439	4.608
Phe	494	494	483	494	494	5.069
Gly	576	576	562	575	576	5.934
His	187	187	182	187	187	1.909
Ile	508	508	488	508	508	5.196
Lys	562	562	555	562	562	5.839
Leu	919	919	884	918	918	9.489
Met	205	205	201	205	205	2.139
Asn	531	531	520	531	531	5.457
Pro	394	393	391	394	394	4.035
Gln	364	364	360	364	364	3.732
Arg	350	350	336	350	350	3.54
Ser	659	660	644	659	660	6.785
Thr	717	716	704	717	717	7.428
Val	780	782	768	781	780	8.084
Trp	110	110	103	110	110	1.118
Tyr	447	447	438	447	447	4.593
Xaa				1		

Molecular weight ranged from 4.449 to 794.057 kDa and isoelectric point (pI) ranged from 4.495 to 9.487

The molecular weight of the CoVid19 proteins ranged from 4.449 (ORF10) to 794.057 (ORF1ab) kDa) (Supplementary Table 5). The molecular weight of the other SARS-CoV-2 proteins were 141.178 (surface glycoprotein), 45.625 (nucleocapsid phosphoprotein), 31.122 (ORF3a), 25.146 (membrane glycoprotein), 13.831 (ORF8), 8.365 (envelope protein), and 7.272 (ORF6) kDa) (Supplementary Table 5). Except for ORF1ab and surface glycoprotein, all other

proteins were found below 50 kDa. The *pI* of SARS-CoV-2 proteome ranged from 4.495 (ORF6) to 9.487 (nucleocapsid phosphoprotein) (Supplementary Table 6). The ORF1ab (5.982), surface glycoprotein (5.906), ORF3a (5.321), ORF8 (5.219) and ORF6 (4.495) were found to have *pI* below seven) (Supplementary Table 6). Analysis of palmitoylation sites in CoVid19 proteins revealed the presence of palmitoylation sites in SARS-CoV-2 proteins (Supplementary Table 7). Co-valent attachment of palmitic acid occurs at the cysteine residue of the protein to increase the hydrophobicity and membrane association (Supplementary Figure 10).

Supplementary Table 5 Molecular weight of SARS-CoV-2 proteins

Proteins	Molecular Weight (KDa) of SARS-CoV-2 proteins from different Countries				
	China	India	Italy	USA	Nepal
ORF1ab	794.0578	794.0719	793.9446	794.0578	794.0578
Surface glycoprotein (S)	141.1785	141.2065	141.1785	141.1785	141.1785
ORF3a	31.12294	31.12294	31.16502	31.12294	31.12294
Envelope protein (E)	8.36504	8.36504	8.36504	8.36504	8.36504
Membrane glycoprotein (M)	25.14662	25.14662	25.14662	25.14662	25.14662
ORF6	7.27254	7.27254	7.27254	7.27254	7.27254
ORF7a	13.74417	13.74417	13.74417	13.74417	13.74417
ORF8	13.83101	13.80493	13.83101	13.80493	13.83101
Nucleocapsid (N)	45.6257	45.6257	45.6257	45.6257	45.6257
ORF10	4.44923	4.44923	4.44923	4.44923	4.44923

Supplementary Table 6 Isoelectric point of SARS-CoV-2 proteins

Proteins	pI of SARS-CoV-2 proteins from different Countries				
	China	India	Italy	USA	Nepal
ORF1ab	5.982	5.982	5.982	5.982	5.982
Surface glycoprotein (S)	5.906	5.906	5.906	5.906	5.906
ORF3a	5.321	5.321	5.321	5.321	5.321
Envelope protein (E)	7.761	7.761	7.761	7.761	7.761
Membrane glycoprotein (M)	9.084	9.084	9.048	9.048	9.048
ORF6	4.495	4.495	4.495	4.495	4.495
ORF7a	7.249	7.249	7.249	7.249	7.249
ORF8	5.219	5.219	5.219	5.219	5.219
Nucleocapsid (N)	9.487	9.487	9.487	9.487	9.487
ORF10	8.302	8.302	8.302	8.302	8.302

Supplementary Table 7 Prediction of palmitoylation sites in SARS-CoV-2 proteins

Proteins	Palmitoylation	Sites	Score
Envelope protein	ILTALRLCAYCCNIV	40	7.195
	ALRLCAYCCNIVNVS	43	18.349
	LRLCAYCCNIVNVSL	44	7.773
Membrane glycoprotein	NA	NA	NA
Nucleocapsid phosphoprotein	NA	NA	NA
ORF1ab	ARAGKASCTLSEQLD	213	15.061
	GHNLAKHCLHVVGPN	1114	11.529
	NSQTSLRCGACIRRP	5340	14.034
ORF3a	IIMRLWLCWKCRSKN	130	4.168
ORF6	NA	NA	NA
ORF7a	ALITLATCELYHYQE	15	23.709
	ELYHYQECVRGTTVL	23	14.058
ORF8	VAAFHQECSLQSCTQ	20	12.122
ORF10	TIYSLLLCRMNSRNY	19	24.543
Surface glycoprotein	LPLVSSQCVNLTTRT	15	23.1
Myristylation			
NA	NA	NA	NA

Discussion

Sequence analysis and similarity study of SARS-CoV-2 (CoVid19) genomes with bat SARS CoVs, MERS CoV, human CoV HKU1 and other revealed that bat SARS CoV and human SARS CoVs (229E German isolate) are not the direct and immediate contributor to the human SARS-CoV-2 (CoVid19) genome. If the genome would have come from either bat SARS CoV or human SARS CoVs 229E, there would be more than 99% of sequence similarity with the direct donor. The rate of mutation of the nucleotides are not so frequent that SARS-CoV-2 (CoVid19) will mutate to such an extent that at short frame of time (a few months) it will result only 80% sequence similarity with bat SARS CoV or human SARS CoV 229E (Table 2). The mutation rate of human genome is 2.5×10^{-8} or 175 mutation per diploid genome per generation.[5] The mutation rate of RNA viral genome ranged from 10^{-6} to 10^{-4} substitution per nucleotide and nucleotide substitution are more common than insertions or deletions.[6] The human SARS CoV 229E genome of German isolate reported long ago in 2003 and it's also showed only 60% sequence similarity with SARS-CoV-2 (CoVid19) (Table 2). However, when sequence similarity study of SARS-CoV-2 was conducted with recent reports of SARS-CoV-2 isolates from China, India, Nepal, and USA, it showed 99% to 100% sequence similarity with each other (excluding SARS CoV 229E German isolate) (Table 2). The phylogenetic tree also did not show any close grouping with the bat SARS CoV (Figure 1). The bat SARS CoV falls in a separate group in the phylogenetic tree and if the SARS-CoV-2 genome would have directly come from bat SARS CoV, they would have certainly grouped with the SARS-CoV-2 genomes (Figure 1). The classic example is that, human SARS CoV 229E of German isolate reported in 2003 fall far distantly. The recent isolates of SARS-CoV-2 of different countries have not undergone significant mutation.

Instead it was observed that, the recent SARS-CoV-2 genomes have undergone some substitutions. The substituted G nucleotide instead of A (Indian isolate), substituted T nucleotide instead of A (Italian isolate), substituted T nucleotide instead of C (Indian and USA isolates), substituted N nucleotide instead of G (Italian isolate), substitution of T nucleotide instead of C (Indian and Nepal isolate), and substituted T nucleotide instead of G (Italian isolates) were the classic examples of SARS-CoV-2 substitution (Supplementary Figure 3). Maximum composite likelihood analysis for pattern nucleotide substitution resulted high rate of transition from T to C nucleotide and a lower rate of transversion (Table 3). However, the transition rate of the genomes of SARS-CoV-2 isolates of countries China, India, Italy, Nepal, and USA was lower than the transition rate of SARS CoVs with bat CoV, MERS CoV, SARS CoV of Canada, bovine CoV, SARS CoV of Germany and others. The time tree analysis also revealed the recent origin of SARS-CoV-2 which date back to 0.00 million years ago, suggesting their evolution from a recent synthetic source (Figure 2). Study reported the shifting of SARS CoV from one host to another.[7] Study also reported about the recombination history of bat SARS CoV of Kenya and German isolate 229E.[8,9] However, our analysis did not result any recombination within the SARS CoV genomes or SARS-CoV-2 genome (Figure 3), suggesting their recent synthetic origin.

The substitution of nucleotides led to the substitution of amino acids in the CDS. In ORF1ab that encode for viral RNA polymerase found to have amino acid P>L (Indian isolate) substitution, L>X (Italian isolate) substitution, and T>I (Indian isolate) substitution. Indian isolate has two substitutions in ORF1ab. The substitution of amino acid P to L in human immune deficiency (HIV) reverse transcriptase (RT) virus led to sensitize RT7 to 10 folds to Nevirapine

antiviral drug.[10] However, the substitution of amino acid T>I show resistant to ganciclovir in human cytomegalovirus.[10] The substitution of amino acid A>V found in surface glycoprotein of Indian isolates. The substitution of amino acid A>V in Zika virus NS2A protein affects viral RNA synthesis and attenuates the virus *in vivo*.[11] Substitution of amino acid G>V was found in ORF3a in Italian SARS-CoV-2 isolate. The substitution of amino acid G>V in *Thermoplasma acidophilum* citrate synthase interfere with the stability and activity of the protein. It also lead to the temperature sensitive altered drug resistance in cytoplasmic loop of the P-glycoprotein.[12,13] In addition, substitution of amino acid G>V lead to delayed folding in type-I pro-collagen protein.[14] ORF8 has amino acid L>S substitution in Indian and USA CoVid19 isolate. The substitution of amino acid L>S induces mecillinam and quinolones resistance.[15,16] The genomic and CDS sequences of the SARS-CoV-2 isolates contained short microsatellite repeats and the presence of microsatellite repeats might favours the substitution and polymorphism in SARS-CoV-2 genome.[17,18] The substitution and recombination study of bat CoV was studied long before and it was reported the coexistence of different genotype in the same bat.[19] However, no such different genotype was observed in the human SARS-CoV-2 till now. Lau et al., (2010) conducted a recombination study of bat corona virus Ro-batCoV HKU9 genome and generated a recombinant bat CoV.[19] However, they have not mentioned what was the possible objective and implication of the generated recombinant bat CoV. The lack of high sequence similarity of SARS-CoV-2 genome with bat and CoV genome proved that, the present SARS-CoV-2 genome did not come from the bat CoV directly. Indeed, the skeleton was sourced from the bat CoV and some synthetic nucleotides were inserted in the bat CoV genome to generate a SARS-CoV-2 genome. Further, human SARS CoV 229E and Chinese SARS CoV (accession: NC_045512.2) had 12 CDS, Canada SARS CoV (accession: NC_004718.3) had 14 CDS whereas the SARS-CoV-2 contain only 10 CDS. It is yet to know why the previous SARS CoV genome contained 12-14 CDS and recent SARS-CoV-2 (CoVid19) genome contained only 10 CD. In addition, the generation of recombinant bat CoV genome by Lau et al.,[19] directly linked towards the generation of recombinant/synthetic CoV genome. This proves that the recent CoVid19 genome might be synthetic in origin.

Proteomic analysis revealed, out of ten SARS-CoV-2 proteins, six of them are have melting temperature (Tm) ranged between 55-65°C whereas ORF6, ORF8, and ORF10 had Tm greater than 65°C. Only the membrane glycoprotein had Tm below 55°C (Supplementary Table 2). If the membrane glycoprotein of SARS-CoV-2 possess Tm less than 55°C, this protein most possibly highly temperature sensitive and this protein can be targeted to destabilize SARS-CoV-2 through application of high temperature. Application of steam through the airways (nose and mouth) has the potential to destabilize the CoVid19 surface glycoprotein and if a person at the early stage of infection receives steam treatment it can be of very useful to reduce the impact of the virus. Chan et al.,[20] reported that the viability of SARS CoV lost at >3Log$_{10}$ at 38°C and relative humidity of greater than 98%. Therefore, the steam application can be a highly viable method to fight SARS-CoV-2 as it will provide high temperature and humidity simultaneously. L-arginine is used to supress the protein aggregation.[21] Therefore, application of saline drips with L-arginine supplement to the SARS-CoV-2 patient may inhibit the aggregation of viral proteins inside the cell thereby lowering the formation of more virus inside the cell. This might be a valuable step towards the suppression of formation of new SARS-CoV-2.

Material and methods

Sequence data

Various corona virus isolates were downloaded from the NCBI database. In total 30 full length corona virus genomes were downloaded. They were bat SARS CoV HKU 3-1 (accession: DQ022305.2), bat SARS CoV WIV1 (accession: KF367457.1), bovine CoV (accession: NC_003045.1), beta CoV from Canada (accession: NC_004718.3), SARS CoVid19 (CoV2) from China (accession: NC_045512.2, MN938384.1, MN975262.1, MN988668.1, MN988669.1, MN996527.1, MN996528.1, MN996529.1, MN996530.1, MN996531.1, MT135041.1, MT135043.1, and MN908947.3), human SARS CoV from Germany (accession: NC_002645.1), SARS CoV2 India (accession: MT050493.1), SARS CoV2 Italy (accession: MT066156.1), MERS CoV (accession: NC_019843.3), SARS CoV2 from Nepal (accession: MT072688.1), beta CoV from the United Kingdom (accession: KC164505.2), and SARS CoV2 of the United States of America (accession: MN985325.1, MN988713.1, MN994467.1, MN994468.1, MN997409.1, MT027063.1 and MT027062.1). The term CoV2 was kept for recently sequence CoVid19 genome originated from the CoVid19 patient.

Analysis of sequence similarity

To find the possible donor of human SARS-CoV-2 from bat CoVs, we aligned the full-length whole genome sequences of SARS-CoV-2 isolates of China, India, Italy, Nepal, and USA with the human SARS CoV isolates of German, bat SARS CoV HKU3-1, bat SARS CoV WIV1, MERS CoV, and SARS CoV2 Wuhan-Hu-1. Sequence alignment was conducted using MUSCLE program (https://www.ebi.ac.uk/Tools/msa/muscle/). We aligned the full-length genome sequence isolates of China, India, Nepal, Italy, and China SARS-CoV-2 genomes to understand the nucleotide similarity and variation among them. There was more than 12 SARS-CoV-2 isolates from China alone. We aligned all the SARS-CoV-2 Chinese isolates to find the variation within the Chinese population. Similarly, there was seven SARS-CoV-2 isolates from the USA. We also aligned all the full-length SARS-CoV-2 genomes of the USA isolates together. The full length CDS and protein sequences were downloaded from the NCBI in fasta format. Multiple sequence alignment of CDS sequences were also conducted using the MUSCLE programme. The protein sequences of the SARS-CoV-2 proteins were aligned using Multalin software to find the substitution in amino acids. The presence of microsatellite markers in the SARS-CoV-2 genome was analysed using the microsatellite repeat finder (http://insilico.ehu.es/mini_tools/microsatellites/). Default parameters were used to find the microsatellite repeats.

Construction of the phylogenetic tree

To construct the phylogenetic tree, the CDS sequences and full-length whole genome sequences of SARS-CoV-2 genomes were aligned using MUSCLE multiple sequence alignment program. The aligned sequence files were converted to MEGA file format using MEGA6 software.[22] Prior to the construction of the phylogenetic tree, a model selection was conducted in MEGA6 software. The phylogenetic tree was constructed using the lowest BIC score of the model selection result. The phylogenetic tree was constructed using the maximum likelihood approach. The statistical parameters used to construct the phylogenetic tree was; model/method, general time reversible model; substitution type, nucleotides; rates among sites,

gamma distributed with invariant sites (G+I); no of discrete gamma parameters, 5; and number of bootstrap replicates, 1000. The codon usage bias and maximum likelihood estimate of substitution was studied using MEGA6 software. The program used to analyse the maximum likelihood substitution was; substitution pattern estimation (ML); model/method, general time reversible model; rates among sites, gamma distributed with invariant sites; number of discrete gamma parameters, 5. The time tree (Reltime ML) was conducted using MEGA6[22].The recombination event study of CoVid19 with other SARS CoVs were analysed using the IcyTree.[23]

CoVid19 proteome analysis

The isoelectric point and molecular weight of the SARS-CoV-2 proteins of the isolates of China, Indian, Italy, Nepal, and USA were calculated using IPC isoelectric point calculator in a Linux based platform.[24] The amino acid composition was also calculated using a Linux based code. The principal component analysis of amino acid composition of the SARS-CoV-2 proteins was conducted using scientific statistical analysis software Past3 (https://folk.uio.no/ohammer/past/). The half-life period of the SARS-CoV-2 proteins was calculated using Protoparam tool (https://web.expasy.org/protparam/).[19] The melting temperature (Tm) of CoVid19 proteins were analysed using protein Tm predictor (http://tm.life.nthu.edu.tw/). Palmitoylation of SARS-CoV-2 proteins was analysed using CSS palm software.[25]

Conclusion

The lack of significant sequence similarity of bat SARS CoV genome with SARS-CoV-2 genome showed the origin of SARS-CoV-2 other than bat SARS or human SARS CoV (German). Most possibly it was a synthetic genome (with bat CoV as a skeleton) as no recombination events was found within or between SARS CoVs. The phylogenetic tree also supported the origin of SARS-CoV-2 other than bat SARS CoV. The time tree analysis also revealed the recent origin of SARS-CoV-2. The publication by Lau et al.,[19] support the finding that laboratory based recombination study of bat SARS CoV was conducted in China to generate recombinant bat SARS CoV.[19] The lack of explanation regarding the application of recombinant bat SARS CoV by Lau et al.,[19] make it doubtful towards the natural origin of SARS-CoV-2. The presence of low Tm of CoVid19 surface glycoprotein might get destabilize by the application of high temperature steam to stop the viral activities.

Conflicts of interest

There is no competing of interest to declare.

References

1. Fan Y, Zhao K, Shi ZL, et al. Bat Coronaviruses in China. *Viruses*. 2019;11(3):210.

2. Drosten C, Günther S, Preiser W, et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N Engl J Med*. 2003;348(20):1967–1976.

3. York A. Novel coronavirus takes flight from bats? *Nat Rev Microbiol*. 2020;18:191.

4. Snijder EJ, Bredenbeek PJ, Dobbe JC, et al. Unique and conserved features of genome and proteome of SARS-coronavirus an early split-off from the coronavirus group 2 lineage. *J Mol Biol*. 2003;331(5):991–1004.

5. Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics*. 2000;15:297–304.

6. Sanjuán R, Nebot MR, Chirico N, et al. Viral Mutation Rates. *J Virol*. 2010;84:9733 LP–9748.

7. Cui J, Han N, Streicker D, et al. Evolutionary relationships between bat coronaviruses and their hosts. *Emerg Infect Dis*. 2007;13:1526–1532.

8. Corman VM, Baldwin HJ, Tateno AF, et al. Evidence for an Ancestral Association of Human Coronavirus 229E with Bats. *J Virol*. 2015;89:11858–11870.

9. Tao Y, Shi M, Chommanard C, et al. Surveillance of Bat Coronaviruses in Kenya Identifies Relatives of Human Coronaviruses NL63 and 229E and Their Recombination History. *J Virol*. 2017;91(5):e01953–16.

10. Dueweke TJ, Pushkarskaya T, Poppe SM, et al. A mutation in reverse transcriptase of bis(heteroaryl)piperazine-resistant human immunodeficiency virus type 1 that confers increased sensitivity to other nonnucleoside inhibitors. *Proc Natl Acad Sci*. 1993;90:4713–4717.

11. Wolf DG, Smith IL, Lee DJ, et al. Mutations in human cytomegalovirus UL97 gene confer clinical resistance to ganciclovir and can be detected directly in patient plasma. *J Clin Invest*. 1995; 95: 257–263.

12. Márquez-Jurado S, Nogales A, Ávila-Pérez G, et al. An Alanine-to-Valine Substitution in the Residue 175 of Zika Virus NS2A Protein Affects Viral RNA Synthesis and Attenuates the Virus *In Vivo. Viruses*. 2018;10(10):547.

13. Loo TW, Clarke DM. Functional consequences of glycine mutations in the predicted cytoplasmic loops of P-glycoprotein. *J Biol Chem*. 1994;269(10):7243–7248.

14. Tsuneyoshi T, Westerhausen A, Constantinou CD, et al. Substitutions for glycine α1-637 and glycine α2-694 of type I procollagen in lethal osteogenesis imperfecta: The conformational strain on the triple helix introduced by a glycine substitution can be transmitted along the helix. *J Biol Chem*. 1991;269(20):5608–15613.

15. Bouloc P, Vinella D, D'Ari R. Leucine and serine induce mecillinam resistance in *Escherichia coli. Mol Gen Genet*. 1992;235:242–246.

16. Ogbolu DO, Daini OA, Ogunledun A, et al. Effects of gyrA and parC mutations in quinolones resistant clinical gram negative bacteria from Nigeria. *African J Biomed Res*. 2012;15:97–104.

17. Brohede J, Ellegren H. Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. *Proc R Soc London Ser B Biol Sci*. 1999;266:825–833.

18. Zhu Y, Strassmann JE, Queller DC. Insertions substitutions and the origin of microsatellites. *Genet Res*. 2000;76:227–236.

19. Lau SP, Poon RS, Wong BL, et al. Coexistence of Different Genotypes in the Same Bat and Serological Characterization of Rousettus Bat Coronavirus HKU9 Belonging to a Novel Betacoronavirus Subgroup. *J Virol*. 2010;84:11385–11394.

20. Chan KH, Peiris JM, Lam SY, et al. The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. *Adv Virol*. 2011;2011:734690.

21. Lange C, Rudolph R. Suppression of Protein Aggregation by L-Arginine. *Curr Pharm Biotechnol*. 2009;10:408–414.

22. Tamura K, Filipski A, Peterson D, et al. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol*. 2013;30:2725–2729.

23. Vaughan TG. IcyTree: rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics*. 2017;33:2392–2394.

24. Kozlowski LP. IPC – Isoelectric Point Calculator. *Biol Direct*. 2016;11:55.

25. Ren J, Wen L, Gao X, et al. CSS-Palm 2.0: an updated software for palmitoylation sites prediction. *Protein Eng Des Sel*. 2008;21:639–644.