Journal of Life Sciences & Biomedicine, vol. 2(75), No 2, p. 58-67 (2020)

Comparative studies of E, M and N structural proteins of SARS-CoV, SARS-CoV-2, pangolin CoV and bat CoV

K.G. Gasimov[1], T.A. Samadova[1], F.K. Abasova [2], I.A. Shahmuradov[1,3*]

[1] *Institute of Biophysics, Azerbaijan National Academy of Sciences, 117 academician Zahid Khalilov Str., Baku AZ1141, Azerbaijan*
[2] *Academician Zarifa Aliyeva Lyceum, H.Javid Ave, 253rd block, Baku AZ 1141, Azerbaijan*
[3] *Institute of Molecular Biology & Biotechnologies, Azerbaijan National Academy of Sciences,11 Izzat Nabiyev Str., Baku AZ 1073, Azerbaijan*
**For correspondence: ilhambaku@gmail.com*

Received 26 October 2020; Received in revised form 25 November 2020; Accepted 30 November 2020

During the last two decades, humanity has been plagued by 3 coronavirus diseases, although the human coronaviruses were discovered over 50 years ago. The latest coronavirus disease discovered in 2019 (COVID-19) is caused by human Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *A question arises:* what could be the reason for such activation of coronaviruses in recent years? To answer this question, at least, it is necessary to clarify (1) the history and origin of these viruses, and (2) molecular mechanisms how they very easily and rapidly enter into host cells and cause multifaceted serious disorders. In this study, we compared the structural proteins E, M and N from SARS-CoV-2, SARS, bat and pangolin CoVs. The most striking fact firstly discovered in this study is that the relative proportion of the synonymous substitution rates in M and N proteins of the SARS-CoV-2 and pangolin CoV are significantly higher than the corresponding characteristics for other CoVs studied. This finding puts several intriguing questions on the emergence and the duration of divergence of the SARS-CoV-2.

Keywords: *Coronavirus, COVID-19, E protein, M protein, N protein, pangolin, bat, origin of the SARS-CoV-2, synonymous mutations, non-synonymous mutations*

INTRODUCTION

Coronaviruses (CoVs), enveloped RNA viruses, cause diseases of wide range in mammals and birds (Fehr and Perlman, 2015; Rabi et al., 2020). In 2002-2003, the highly pathogenic Severe Acute Respiratory Syndrome coronavirus, SARS-CoV, was discovered in China. Later (2012), the Middle Eastern Respiratory Syndrome Coronavirus (MERS-CoV) emerged in Kingdom of Saudi Arabia. At last, in December 2019, a new and extremely dangerous disease, COVID-19, associated with the most known pathogenic SARS coronavirus, SARS-CoV-2, was fixed in China (Shi and Hu, 2020). To date (November 19, 2020), there have been about 56,000,000 confirmed cases of COVID-19, including over 1,344,000 deaths (https://www.who.int/emergencies/diseases/novel-coronavirus-2019). Thus, during last two decades, the humanity has been plagued by 3 coronavirus diseases, although the human coronaviruses were discovered over 50 years ago (McIntosh et al., 1967; McIntosh, 1974). Therefore, a natural question arises: what could be the reason for such activation of coronaviruses in recent years? To answer this question, it is necessary to determine the approximate date and origin of the virus in the human body. Moreover, if we understand the events that led to the emergence of human coronaviruses, we can also predict and prevent new pandemics.

Bats are currently considered as one of the potential natural reservoirs of various viruses, including SARS-CoV and MERS-CoV (for a review see: Cui et al., 2019). Indeed, studies indicate that many coronaviruses are capable of interspecies transmission (Tang et al., 2015). In partic-

doi.org/10.29228/jlsb.62

Available online 25 December 2020

ular, some bat coronaviruses and SARS-CoV can use the same receptor to enter cells (Hu et al., 2015; Menachery et al., 2015). However, recent findings indicate that SARS-CoV-2-like CoVs might originate from pangolin species (Lopes et al., 2020; Malaiyan et al., 2020; Tang et al., 2020; Zhang et al., 2020).

Coronaviruses have a positive-sense RNA genome of ~30 kb, with a 5′-cap and a 3′ poly (A)-tail structure. This structure allows RNA to serve as a direct mRNA for synthesis of the viral polypeptides (Fehr and Perlman, 2015; Hu et al., 2015). In particular, the SARS-CoV-2 genome (29,880 bp) encodes four main structural, spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins, and nonstructural (3-chymotrypsin like protease, papain-like protease, and RNA-dependent RNA polymerase) proteins (for review see: Huang et al., 2020).

The SARS-CoV-2 entrance into the host cell is initiated by interactions between the ~150 kDa S protein and its receptor, the angiotensin-converting enzyme 2 (ACE2). This protein is composed of the N-terminal S1 (13-685 aa) and the C-terminal S2 (686-1270 aa) subunits gained by a cleavage of the primary S protein by a host cell furin-like protease. The host ACE2 is recognized and bound by the S1 subunit, while the S2 is required mainly for a fusion of viral and host cell membranes. Thus, protein S appears to be an important determinant of CoV pathogenesis and resistance to infection in the body.

The small (~8-12 kDa) and probably transmembrane E protein is found in small quantities within the virion. Although the E proteins have been found to vary greatly in different CoV groups, they share a common architecture. In contrast to other structural proteins, recombinant viruses lacking the E protein are not always lethal. The E protein is mostly involved in assembly and release of the virus, but is also required for pathogenesis (Fehr and Perlman, 2015; (Bianchi et al., 2020; for a review see: Satarker et al., 2020).

The N protein (~49.5 kDa) is an important antigen for CoV, which participates in RNA package and virus particle release (Zeng et al., 2020). N protein is second most abundant viral protein, and is expressed during the early stages of infection. It is composed of two separate domains, an N-terminal domain (NTD) and a C-terminal domain (CTD), both of which are capable of binding RNA *in vitro*. However, the optimal RNA binding function of the N protein is supposed to be required for both domains. This protein helps to enter the host cell and interact with cellular processes after the virus fusion. (Huang et al., 2004; V'kovski et al., 2019). The SARS-CoV-2 and SARS-Cov N protein sequence shows about 90% similarity (Gralinski and Menachery, 2020). N protein of SARS-CoV promotes the activation of cyclooxygenase-2 (COX-2) and causes inflammation in the lungs (Yan et al., 2006). It also participates in the inhibition of a phosphorylation of the B23 protein, which is involved in the development of the cell cycle (Zeng et al., 2008), as well as in inhibition of the viral proteins degradation (Wang et al., 2010). Moreover, N protein restricts immune responses in the body against the viral infections via inhibition of the type I interferon (Lu et al., 2011).

The M protein (~25-30 kDa) is the most abundant structural protein in the virion (Alsaadi and Jones, 2019). It contains 3 transmembrane domains and is required for the shaping and budding processes of CoVs (Bianchi, et al., 2020). Most M proteins do not contain a signal sequence, although they are co-translationally inserted in the endoplasmic reticulum membrane (Fehr and Perlman, 2015). Structural analysis of the M protein indicates its existence in two, long and compact, forms. This protein inhibits the Nuclear Factor kappa-light-chain-enhancer of activated B cells through interactions with I Kappa B Kinase and reduces levels of Cyclooxygenase 2, thus enhancing the proliferation of the viral pathogen (Fang et al., 2007). At last, the protein is known to be involved in the activation of beta-interferons (Satarker et al., 2020).

However, we are too far from understanding molecular mechanisms determining their host range and pathogenesis rate, supposed harmful side effects in the host organisms. In this sense, a comparative exploration of SARS-CoV-2, SARS-CoV and other related coronavirus genomes from human, bat and other species seems to be one of most efficient ways in understanding genetic bases of the CoV problem. In particular, the whole-genome sequencing and analysis data on SARS-CoV-2 from different populations are recently emerging (Munnink et al., 2020; Meredith et al., 2020).

In this study, the E, M and N proteins from SARS-CoV-2, SARS-Cov, bat CoV and pangolin CoV are compared. Below, we present and discuss results of these studies.

MATERIALS AND METHODS

For analysis, both CDS and protein sequences of the E, M and N proteins from the human, pangolin and bat CoVs, including SARS-CoV-2 (GenBank accession MN997409.1), SARS-CoV E (NC_004718.3), the pangolin CoV (MT040335.1), 9 strains of the bat CoVs (strain 273/2005: GenBank accession: ABG47063.1; 279/2005: ABG47072.1; Italy/206679-3/2010: AZF86133.1; Italy/206645-41/2011: AZF86121.1; Italy/3398-19/2015: AZF86127.1; Rm1/2004: ABD75325.1; Rp3/2004: AAZ67055.1; HKU9: YP_001039974.1; Vs-CoV-1: BBJ36014.1) were used.

A comparison of CDS and protein sequences was done by BLAST tool (Altschul et al., 1997). A multiple alignment of CDS and protein sequences, as well as the construction of the phylogenetic trees was performed by the Clustal Omega tool (Sievers and Higgins, 2014; Sievers and Higgins, 2018).

To investigate the statistical characteristics of variations, such as identities, synonymous and nonsynonymous substitutions, as well as insertions/deletions (Indels), as a new tool, MUTAN-2 was developed by I. Shahmuradov (unpublished). An output of the pairwise alignment of protein sequences by the Clustal Omega (in the FASTA format) and corresponding query CDS sequences serve as a source (input) data for this tool.

RESULTS AND DISCUSSION

Initially, using the BLAST and Clustal Omega tools, we compared E, M and N protein sequences from the human SARS-CoV-2 and SARS-CoV, as well the pangolin CoV and 9 strains of the bat CoV (see: Materials and Methods). Results of these comparisons are illustrated in Table 1, 2 and 3, as well as in Fig. 1, 2 and 3. Proteins E and M, as well as human and pangolin CoV proteins show significant (88% or higher) similarity.

However, only 4 (out of 9) strains (273/2005, 279/2005: Rm1/2004 and Rp3/2004) of bat CoV were found to have significant (88% or higher) similarity to the corresponding human and pangolin CoV proteins. The same results were obtained for S proteins (a paper on comparative studies of the CoV S-proteins was recently submitted elsewhere).

Table 1. Percent identity matrix for the *E* proteins from the human SARS-CoV and SARS-CoV-2, pangolin CoV and 9 strains of the bat CoV

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1. Bat (Italy-1)		100.0	70.67	16.00	20.00	16.67	16.67	16.44	16.44	17.81	17.81	17.81
2. Bat (Italy-3)	100.0		70.67	16.00	20.00	16.67	16.67	16.44	16.44	17.81	17.81	17.81
3. Bat (Italy-2)	70.67	70.67		18.67	22.67	15.28	15.28	15.07	15.07	16.44	16.44	16.44
4. Bat (HKU9)	16.00	16.00	18.67		20.25	25.68	25.68	26.67	26.67	26.67	26.67	26.67
5. Bat (Vs)	20.00	20.00	22.67	20.25		40.00	40.00	40.79	40.79	40.79	40.79	40.79
6. CoV-2	16.67	16.67	15.28	25.68	40.00		**100.0**	**96.00**	**96.00**	**94.67**	**94.67**	**94.67**
7. Pangolin	16.67	16.67	15.28	25.68	40.00	**100.0**		**96.00**	**96.00**	**94.67**	**94.67**	**94.67**
8. SARS	16.44	16.44	15.07	26.67	40.79	**96.00**	**96.00**		**100.0**	**98.68**	**98.68**	**96.05**
9. Bat (Rp3)	16.44	16.44	15.07	26.67	40.79	**96.00**	**96.00**	**100.0**		**98.68**	**98.68**	**96.05**
10.Bat (279)	17.81	17.81	16.44	26.67	40.79	**94.67**	**94.67**	**98.68**	**98.68**		**100.0**	**96.05**
11.Bat (Rm1)	17.81	17.81	16.44	26.67	40.79	**94.67**	**94.67**	**98.68**	**98.68**	**100.0**		**96.05**
12.Bat (273)	17.81	17.81	16.44	26.67	40.79	**94.67**	**94.67**	**96.05**	**96.05**	**96.05**	**96.05**	

Hereinafter, the following abbreviations are used: “Italy-1” for the Italy/206679-3/2010, “Italy-2” for the Italy/206645-41/2011, “Italy-3” for the Italy/3398-19/2015, “Vs” for the Vs-CoV-1, “Rm1”, “Rp3” for the Rp3/2004, “273” for the 273/2005 and “279” for the 279/2005. A group of CoVs with the significant similarity of *E*-proteins are highlighted in grey. Here, as well as in Table 2 and 3, the SARS-Cov-2 and pangolin CoV similarity is marked in pink, similarity between SARS CoV and 4 bat CoVs (strains Rp3, 279, Rm1 and 273) is highlighted in red.

Table 2. Percent identity matrix for the *M* proteins from the human SARS-CoV, pangolin CoV and 9 strains of the bat CoV

	1	2	3	4	5	6	7	8	9	10	11	12
1. Bat (Italy-1)		100.0	88.11	28.05	27.60	29.55	29.55	29.09	29.09	29.09	31.22	32.72
2. Bat (Italy-3)	100.0		88.11	28.05	27.60	29.55	29.55	29.09	29.09	29.09	31.22	32.72
3. Bat (Italy-2)	88.11	88.11		29.09	29.55	30.45	30.45	30.45	30.00	30.00	31.96	35.94
4. CoV-2	28.05	28.05	29.09		**98.20**	**90.50**	**90.50**	**89.59**	**89.14**	**89.59**	38.91	40.37
5. Pangolin	27.60	27.60	29.55	**98.20**		**90.95**	**90.95**	**90.50**	**89.59**	**90.05**	38.91	39.91
6. Bat (279)	29.55	29.55	30.45	**90.50**	**90.95**		**100.0**	**97.29**	**95.93**	**97.29**	40.45	40.83
7. Bat (Rm1)	29.55	29.55	30.45	**90.50**	**90.95**	**100.0**		**97.29**	**95.93**	**97.29**	40.45	40.83
8. SARS	29.09	29.09	30.45	**89.59**	**90.50**	**97.29**	**97.29**		**97.74**	**97.29**	40.91	41.28
9. Bat (273)	29.09	29.09	30.00	**89.14**	**89.59**	**95.93**	**95.93**	**97.74**		**98.64**	40.91	40.83
10. Bat (Rp3)	29.09	29.09	30.00	**89.59**	**90.05**	**97.29**	**97.29**	**97.29**	**98.64**		40.91	40.83
11. Bat (HKU9)	31.22	31.22	31.96	38.91	38.91	40.45	40.45	40.91	40.91	40.91		41.28
12. Bat (Vs)	32.72	32.72	35.94	40.37	39.91	40.83	40.83	41.28	40.83	40.83	41.28	

A group of CoVs with the significant similarity of *M*-proteins are highlighted in grey.

Table 3. Percent identity matrix for the *N* protein from the human SARS-CoV, pangolin CoV and 9 strains of the bat CoV

	1	2	3	4	5	6	7	8	9	10	11	12
1. Bat (Italy-1)		98.38	64.11	23.69	27.96	27.90	27.27	27.27	27.59	27.59	27.27	27.90
2. Bat (Italy-3)	98.38		64.11	23.97	27.66	27.59	26.65	26.65	27.27	27.27	26.96	27.59
3. Bat (Italy-2)	64.11	64.11		25.07	29.97	30.37	29.14	29.45	30.06	29.45	29.45	29.75
4. Bat (HKU9)	23.69	23.97	25.07		39.11	44.30	44.92	44.92	44.70	44.81	43.04	43.77
5. Bat (Vs)	27.96	27.66	29.97	39.11		47.85	48.48	48.48	48.23	48.35	48.73	48.98
6. Bat (273)	27.90	27.59	30.37	44.30	47.85		**95.71**	**96.19**	**96.44**	**96.67**	**88.78**	**88.97**
7. Bat (279)	27.27	26.65	29.14	44.92	48.48	**95.71**		**99.52**	**97.14**	**97.38**	**88.76**	**89.18**
8. Bat (Rm1)	27.27	26.65	29.45	44.92	48.48	**96.19**	**99.52**		**97.62**	**97.62**	**89.00**	**89.42**
9. SARS	27.59	27.27	30.06	44.70	48.23	**96.44**	**97.14**	**97.62**		**98.10**	**89.74**	**89.93**
10. Bat (Rp3)	27.59	27.27	29.45	44.81	48.35	**96.67**	**97.38**	**97.62**	**98.10**		**89.26**	**89.93**
11. CoV-2	27.27	26.96	29.45	43.04	48.73	**88.78**	**88.76**	**89.00**	**89.74**	**89.26**		**93.76**
12. Pangolin	27.90	27.59	29.75	43.77	48.98	**88.97**	**89.18**	**89.42**	**89.93**	**89.93**	**93.76**	

A group of CoVs with the significant similarity of *N*-proteins are highlighted in grey.

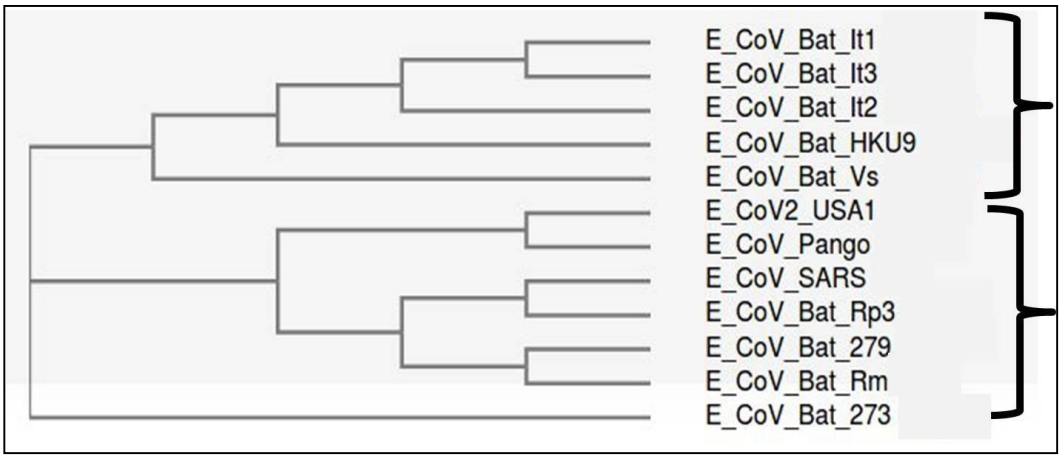

Fig. 1. A phylogenetic tree constructed for on the basis of comparison of E-proteins from 12 CoVs of different species/strains.

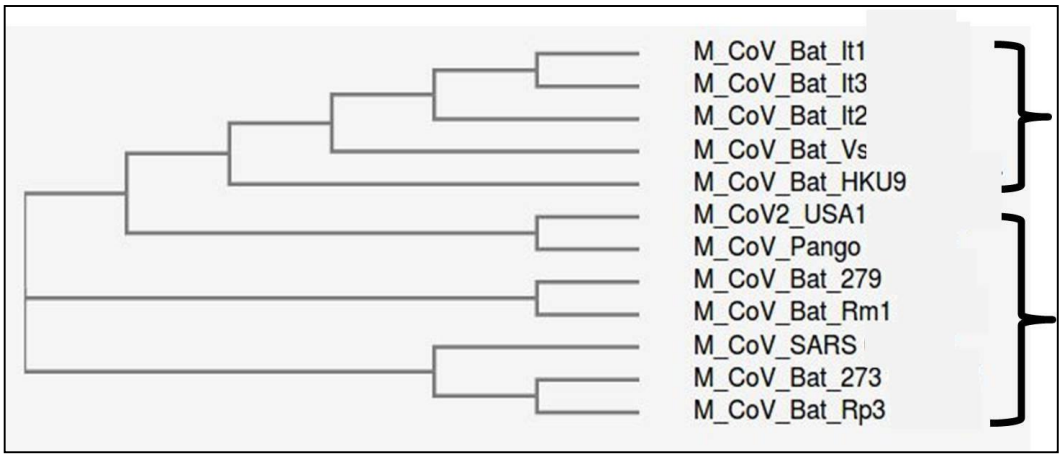

Fig. 2. A phylogenetic tree constructed for on the basis of comparison of M-proteins from 12 CoVs of different species/strains.

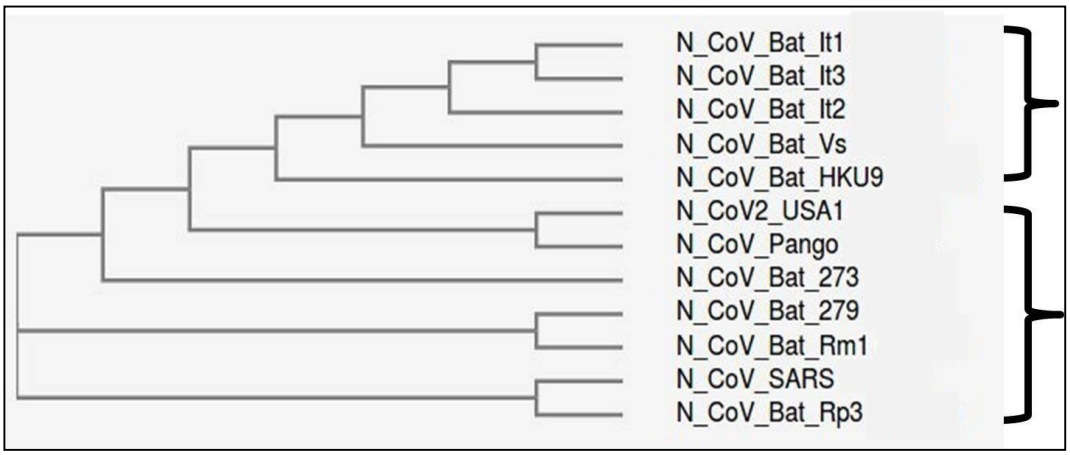

Fig. 3. Phylogenetic tree based on comparison of E-proteins from 12 CoVs of different species/strains.

These findings indicate that at least in bats, various sub-classes of CoVs exist. The different CoV strains in bat have probably been diverged for a long time. Moreover, a significant difference for the similarity level of E, M and N proteins from 4 bat CoV strains is not observed. Taking into account these two facts, for further comparative studies on E, M and N proteins only the following CoVs were selected: SARS-CoV-2, SARS-CoV, pangolin CoV and strain Rp3 of the bat CoV.

First, in each group (E, M and N) of the 4 selected coronaviruses, the proteins were aligned using the Clustal Omega tool. Then, the results of these alignments were analyzed by the MUTAN-2 program. Results of the analysis are summarized in Table 4.

The first remarkable result of these studies is that pangolin CoV is mostly closer to the human SARs-CoV-2 in terms of the inter-species similarity of the E, M and N proteins. Thus, while the similarity between E, M and N proteins from SARS-CoV-2 and pangolin virus is about 100%, 98% and 93%, respectively, these figures for comparisons of SARS-CoV-2 *vs* SARS-CoV and SARS-CoV-2 *vs* bat CoV are 96%, 89% and 89%, respectively. The same picture was observed in the inter-species comparison of S proteins (data not shown). It should be noted that this result is fully consistent with the results of studies recently reported (Lopes et al., 2020; Malaiyan et al., 2020; Zhang et al., 2020).

Compared to proteins M and N, Protein E from human SARS-CoV-2 and SARS-CoV, pangolin CoV, and four strains of bat CoV are almost conserved, although there are also bat CoV strains with significant differences (60-85%; see Table 1). These observations may indicate that while the E proteins are important for the CoV envelope forming, they are not involved in a definition of the host range and pathogenesis of CoVs.

Table 4. The similarity details of the CDS and amino acid sequences of the E, M and N proteins from SARS-COV-2, SARS-CoV, pangolin CoV and bat CoV

	Conservation level: CoV-2 vs Pangolin CoV	**Conservation level: CoV-2 vs SARS CoV**	**Conservation level: CoV-2 vs Bat CoV**
E protein:			
Identities (amino acids)	75 (out of 75), 100.0%	72 (75), 96.0%	72 (75), 96.0%
Identical codons (CDS)	70 (out of 75), 93.33%	65 (75), 86.67%	66/ (75), 88.0%
Synonymous substitutions (CDS)	5 (out of 5), 100.0%	7 (10), 70.0%	6 (9), 66.67%
Non-synonymous substitutions (CDS)	0	3 (10), 30.0%	3 (9), 33.33%
Indels (amino acids)	0	1 (75), 1.33%	1 (75), 1.33%
M protein:			
Identities (amino acids)	218 (222), 98.2%	198 (222), 89.19%	198 (222), 89.19%
Identical codons (CDS)	169 (222), 76.13%	141 (222), 63.51%	3 (193), 1.55%
Synonymous substitutions (CDS)	49 (53), 92.45%	57 (80), 71.25%	59 (82), 71.95%
Non-synonymous substitutions (CDS)	4 (53), 7.55%	23 (80), 28.75%	23 (82), 28.05%
Indels (CDS)	0 (193)	1 (222), 0.45%	1 (222), 0.45%
N protein:			
Identities (amino acids)	391 (419), 93.32%	376 (419), 89.74%	374 (419), 89.26%
Identical codons (CDS)	325 (419), 77.57%	295 (419), 70.41%	291 (419), 69.45%
Synonymous substitutions (CDS)	66 (92), 71.74%	81 (124), 65.32%	83 (128), 64.84%
Non-synonymous substitutions (CDS)	26 (92), 28.26%	43 (124), 34.68%	45 (128), 35.16%
Indels (CDS)	2 (419), 0.48%	3 (419), 0.72%	2 (419), 0.48%

[1]Most significant variations in the amino acid and codon compositions are marked in grey.

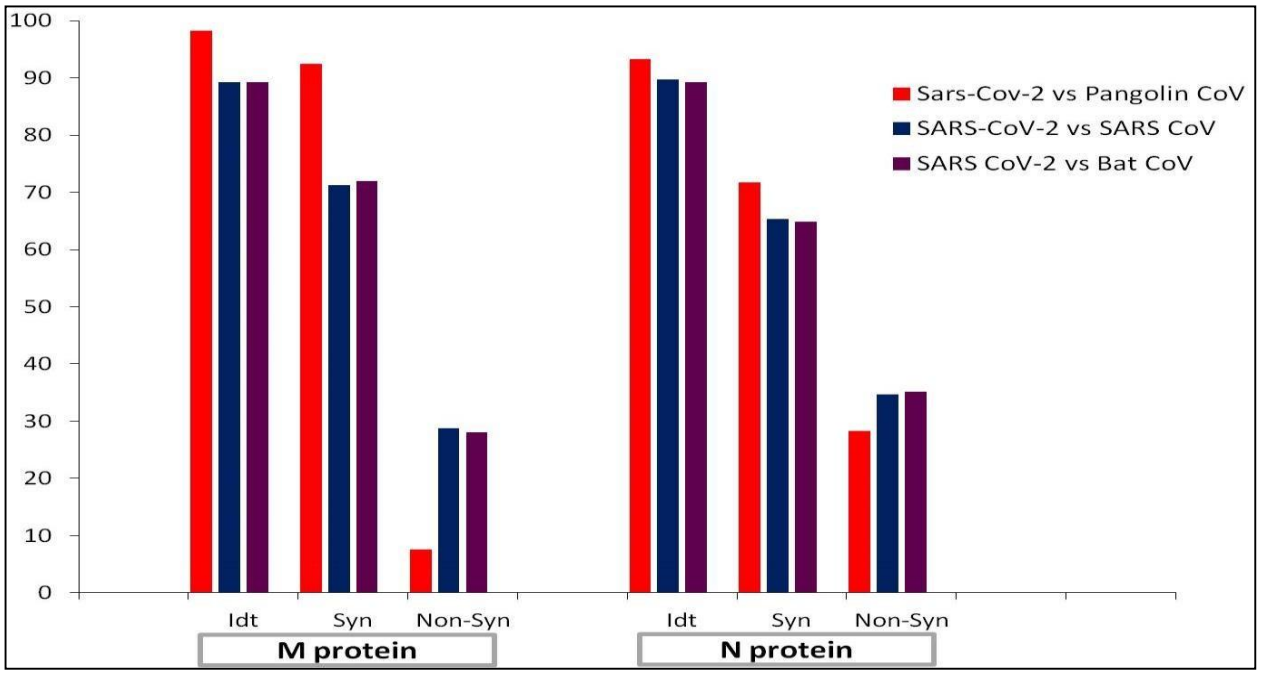

Fig. 4. Graphical representation of differences in identities, synonymous substitutions and nonsynonymous substitutions within M and N proteins between the human SARS-CoV-2 and SARS-CoV, the pangolin CoV and the bat CoV, strain Rp3. Idt – identities, Syn – synonymous mutations, Non-Syn – non-synonymous mutations.

However, the most interesting result was obtained in the comparative studies of the synonymous and non-synonymous substitutions rates in M and N proteins. Thus, out of 53 substitutions in M proteins from the SARS-CoV-2 and pangolin CoV, 49 changes (92.45%) were due to the synonymous substitutions. For the SARS-CoV-2 *vs* ARS-CoV and SARS-CoV *vs* bat CoV comparisons, this rate was significantly lower: 71.25% (57 out of 80) and 71.95% (59 out of 82), respectively. For N proteins, these characteristics were 71.74% (66/92), 65.32% (81/124) and 64.84% (83/128, respectively (Fig. 4; see also Table 4). In particular, these findings suggest that a significant role in almost identity (98.2% similarity) of M proteins in the SARS-CoV-2 and the pangolin CoV belong to the synonymous substitutions. Taking into account these findings and our current knowledge of the key role of M proteins in integration of CoV into the host cell, as well as our recent results on conservation of S proteins in human, pangolin and bat CoVs (unpublished), we suppose that the SARS-CoV and SARS-CoV-2 have some bat CoV and pangolin CoV origin, respectively (Fig. 5).

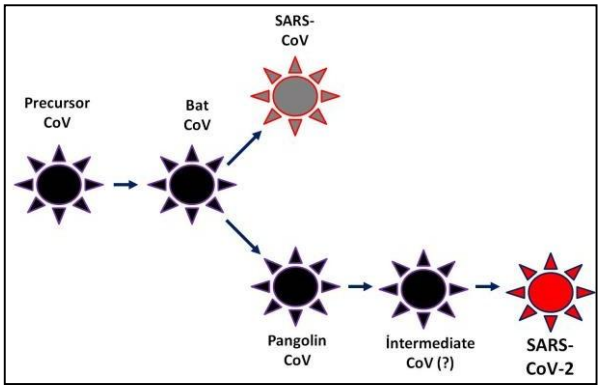

Figure 5. A hypothetical path of evolutionary events resulted in the SARS-CoV-2 and SARS-CoV from a precursor CoV of the unknown origin.

Menachery and colleagues (2015) studied the disease potential for SARS-like virus, SHC014-CoV, from populations of Chinese horseshoe bats. To do this, they developed a chimeric virus and injected it into the mouse backbone. The experimental results showed that the chimeric virus is able to efficiently (i) use orthologues of the SARS receptor for ACE2), (ii) replicate in primary cells of the human respiratory tract, and (iii) achieve *in vitro* titers close to epidemic strains of SARS-CoV. In addition, it has been demonstrated *in vivo* that a recombinant virus can replicate in the lungs of mice with significant pathogenesis using a novel spike protein. These results suggest that full-length recombinant viruses could potentially appear in humans.

It should be noted that our suggestion on the pangolin origin of the SARS-CoV-2 is fully consistent with the results of studies recently reported (Lopes et al., 2020; Malaiyan et al., 2020; Tang et al., 2020; Zhang et al., 2020). But, we understand that the available facts do not preclude a definitive answer to these questions.

CONCLUDING REMARKS

The question on the direct origin of the highly pathogenic SARS-CoV-2 still remains to be answered, although the pangolin origin of this virus seems to be most attractive hypothesis. Moreover, an existence of unknown intermediate organisms in transfer of this virus to humans cannot be excluded.

The question of whether there was human intervention in creation of this virus also remains open. Of course, we are not saying that any malicious people have exposed humanity to a terrible disease like COVID-19 through this coronavirus. However, the possibility of creating a new form of coronavirus to test a scientific idea in a laboratory and then infecting humans as a result of someone's usual negligence, cannot be ruled out.

REFERENCES

Alsaadi E.A., Jones I.M. (2019) Membrane binding proteins of coronaviruses. *Future Virol*., **14:** 275-286.

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25(17):** 3389-3402.

Bianchi M., Benvenuto D., Giovanetti M., Angeletti S., Ciccizzi M., Rascarella S. (2020) Sars-CoV-2 Envelope and Membrane Proteins: Structural Differences Linked to Virus Charac-

teristics? *Bio.Med. Res. International*, **2020:** 4389089. doi.org/10.1155/2020:4389089.

Cardenas-Conejo Y., Linan-Rico A., Garcia-Rodruez D.A., Centeno-Leija S., Serrano-Posada H. (2020) An exclusive 42 amino acid signature in pp1ab protein provides insights into the evolutive history of the 2019 novel human-pathogenic coronavirus (SARS-CoV2). *J. Med. Virol.*, **92(6):** 688-692.

Coutard B., Valle C., de Lamballerie X., Canard B., Seidah N.G., Decroly E. (2020) The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res*., **176**: 104742, doi: 10.1016/j.antiviral.2020.104742.

Cui J., Li F., Shi Z.L. (2019) Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol*., **17(3):** 181-192.

Damas J., Hughes G.M., Keough K.C., Painter C.A., Persky N.S., Corbo M., Hiller M., Koepfli K.P., Pfenning A.R., Zhao H., Genereux D.P., Swofford R., Pollard K.S., Ryder O.A., Nweeia M.T., Lindblad-Toh K., Teeling E.C., Karlsson E.K., Lewin H.A. (2020) Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc. Natl. Acad. Sci. U.S.A.*, **117(36):** 22311-22322.

Fang X., Gao J., Zheng H., Li B., Kong L., Zhang Y., Wang W., Zeng Y., Ye L. (2007) The membrane protein of SARS-CoV suppresses NF-kB activation. *J. Med. Virol.*, **79(10):** 1431-1439.

Fehr A.R., Perlman S. (2015) Coronaviruses: an overview of their replication and pathogenesis. *Methods. Mol. Biol.*, **1282:** 1-23. doi: 10.1007/978-1-4939-2438-7_1.

Gralinski L.E., Menachery V.D. (2019) Return of the coronavirus: 2019-nCoV. *Viruses*, **12:** 1-8.

Hoffmann M., Kleine-Weber H., Schroeder S., Kruger N., Herrler T., Erichsen S., Schiergens T.S., Herrler G., Wu N.-H., Nitsche A. (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, **181(2):** 271-280.

Hu B., Ge X., Wang L.-F., Shi Z. (2015) Bat origin of human coronaviruses. *Virology J.*, **12:** 221. doi: 10.1186/s12985-015-0422-1.

Huang Q., Yu L., Petros A.M., Gunasekera A., Liu Z., Xu N., Hajduk P., Mack J., Fesik S.W., Olejniczak E.T. (2004) Structure of the N-terminal RNAbinding domain of the SARS CoV nucleocapsid protein. *Biochemistry*, **43:** 6059-6063.

Huang Y., Yang C., Xu X.-F., Xum W., Liu S.-W. (2020) Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica*, **41(9):** 1141-1149.

Lopes L.P., Cardillo G.-M., Paiva P.B. (2020) Molecular evolution and phylogenetic analysis of SARS-CoV-2 and hosts ACE2 protein suggest Malayan pangolin as intermediary host. *Brazilian J. of Microbiology*, **June 26:** 1-7. doi: 10.1007/s42770-020-00321-1.

Lu X., Pan J., Tao J., Guo D. et al. (2011) SARS-CoV nucleocapsid protein antagonizes IFN-b response by targeting initial step of IFN-b induction pathway, and its C-terminal region is critical for the antagonism. *Virus Genes*, **42(1):** 37-45.

Malaiyan J., Arumugam S., Mohan K., Radhakrishnan G.G. (2020) An update on the origin of SARS- CoV- 2: Despite closest identity, bat (RaTG13) and pangolin derived coronaviruses varied in the critical binding site and O- linked glycan residues. *Med. Virol.*, **July 7**: 1-7. doi:10.1002/jmv.26261.

McIntosh K., Dees J.H., Becker W.B., Kapikian A.Z., Chanock R.M. (1967) Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci. USA*, **57(4):** 933-940.

McIntosh K. (1974) Coronaviruses: A comparative review. In: W.Arber, R.Haas, W.Henle et al. (eds.). *Current Topics in Microbiology and Immunology/Ergebnisse der Mikrobiologie und Immunitätsforschung*. Berlin, Heidelberg: Springer, p. 85-129.

Menachery V., Yount B., Debbink K., Agnihothram S., Gralinski L.E., Plante J.A., Graham R.L., Scobey T., Ge X.-Y., Donaldson E.F., Randell S.H., Lanzavecchia A., Marasco W.A., Shi Z.-L., Baric R.S. (2015) A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.*, **21(12):** 1508-1513.

Meredith L.W., Hamilton W.L., Warne B., Houldcroft C.J., Hosmillo M., Jahun A.S., Curran M.D., Parmar S., Caller L.G., Caddy

S.L., Khokhar F.A., Yakovleva A., Hall G., Feltwell T., Forrest S., Sridhar S., Weekes M.P., Baker S., Brown N., Moore E., Popay A., Roddick I., Reacher M., Gouliouris T., Peacock S.J., Dougan G., Török M.E., Goodfellow I. (2020) Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.*, **July 14**: doi: 10.1016/S1473-3099(20)30562-4.

Munnink B.B., Nieuwenhuijse D.F., Stein M., O'Toole A., Haverkate M., Mollers M., Kamga S.K., Schapendonk C., Pronk M., Lexmond P., van der Linden A., Bestebroer T., Chestakova I., Overmars R.J., van Nieuwkoop S., Molenkamp R., van der Eijk A.A., Kessel C.G., Vennema H., Meijer A., Rambaut A., van Dissel J., Sikkema R.S., Timen A., Koopmans M., Dutch-Covid-19 response team (2020). Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nature Medicine*, **26(9):** 1405-1410.

Rabi F.A., Al Zoubi M.S., Kasasbeh G.A., Salameh D.M., Al-Nasser A.D. (2020) SARS-CoV-2 and coronavirus disease 2019: What we know so far. *Pathogens*, **9(3)**: 231. doi: 10.3390/pathogens9030231.

Ruch T.R., Machamer C.E. (2011) The hydrophobic domain of infectious bronchitis virus E protein alters the host secretory pathway and is important for release of infectious virus. *J. Virol*., **85:** 675-685.

Satarker S., Nampoothiri M. (2020) Structural proteins in severe acute respiratory syndrome coronavirus-2. *Arch. Med. Res.*, **51(6):** 482-491.

Shi Z., Hu Z. (2020) The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur. J. Clin. Microbiol. Infect. Dis.*, **39(9):** 1629-1635.

Sievers F., Higgins D.G. (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.*, **1079:** 105-116. doi: 10.1007/978-1-62703-646-7_6.

Sievers F., Higgins D.G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.*, **27(1):** 135-145.

Tang Q., Song Y., Shi M. (2015) Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. *Scientific Reports*, **5:** 17155, doi: 10.1038/srep17155.

Tang X., Wu C., Li X., Song Y., Yao X., Wu X., Duan Y., Zhang H., Wang Y., Qian Z. (2020) On the origin and continuing evolution of SARS-CoV-2. *National Science Review*, **7(6):** 1012-1023.

Tian X., Li C., Huang A., Xia S., Lu S., Shi Z., Lu L., Jiang S., Yang Z., Wu Y. (2020) Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg. Microbes Infect*., **9(1):** 382-385.

V'kovski P., Gerber M., Kelly J., Pfaender S., Ebert N., Lagache S.B., Simillion C., Portmann J., Stalder H., Gaschen V., Bruggmann R., Stoffel M.H., Heller M., Dijkman R., Volker T.V. (2019) Determination of host proteins composing the microenviron-ment of coronavirus replicase complexes by proximity-labeling. *Elife*, **8:** e42037.

Wang H., Li X., Li T., Zhang S., Wang L., Wu X., Liu J. (2020) The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur. J. Clin. Microbiol. Infect. Dis*., **39(9):** 1629-1635.

Wang Q., Li C., Zhang Q., Wang T., Li J., Guan W., Yu J., Liang M., Li D. (2010) Interactions of SARS Coronavirus nucleocapsid protein with the host cell proteasome subunit p42. *Virol. J.*, **7:** 99. doi: 10.1186/1743-422X-7-99.

Yan X., Hao Q., Mu Y., Timani K.A., Ye L, Zhu Y., Wu J. (2006) Nucleocapsid protein of SARS-CoV activates the expression of cyclooxygenase-2 by binding directly to regulatory elements for nuclear factor-kappa B and CCAAT/enhancer binding protein. *Int. J. Biochem. Cell Biol.*, **38(8):**1417-1428.

Zeng W., Liu G., Ma H., Zhao D., Yang Y, Liu M., Mohammed A., Zhao C., Yang Y., Xie J., Ding C., Ma X., Weng J., Gao Y., He H., Jin T. (2020) Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem. Biophys. Res. Commun.*, **527(3):** 618-623

Zeng Y., Ye L., Zhu S., Zheng H., Zhao P., Cai W., Su L., She Y., Wu Z. (2008) The nucleocapsid protein of SARS associated coronavirus inhibits B23 phosphorylation. *Biochem. Biophys. Res. Commun.*, **369(2):** 287-291.

Zhang T., Wu Q., Zhang Z. (2020) Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Current Biology*, **30(7):** 1346-1351.

SARS-CoV, SARS-CoV-2, kələzin və yarasanın CoV viruslarının E, M və N struktur zülallarının müqayisəli tədqiqi

K.Q. Qasımov[1], T.A. Səmədova[1], F.K. Abbasova [2], I.Ə. Şahmuradov[1,3*]

[1] AMEA-nın Biofizika İnstitutu, Bakı, Azərbaycan
***2** Akademik Zərifə Əliyeva adına lisey, Bakı, Azərbaycan*
[3] AMEA-nın Molekulyar Biologiya və Biotexnologiyalar İnstitutu, Bakı, Azərbaycan

Son iyirmi il ərzində bəşəriyyət 3 koronavirus xəstəliyinə düçar olmuşdur, halbuki insan koronavirusları 50 il əvvəl aşkar edilmişdir. 2019-cu ildə kəşf olunan son koronavirus xəstəliyinə (COVID-19) insanın şiddətli kəskin tənəffüs sindromu koronavirusu-2 (SARS-CoV-2) səbəb olur. Sual doğur: son illərdə koronavirusların bu cür aktivləşməsinin səbəbi nə ola bilər? Bu suala cavab vermək üçün, heç olmasa, (1) bu virusların tarixini və mənşəyini və (2) çox asanlıqla və sürətlə sahib hüceyrələrinə daxil olaraq çoxşaxəli ciddi pozğunluqlar törətməsinin molekulyar mexanizmləri aydınlaşdırmaq lazımdır. Bu işdə SARS-CoV-2, SARS, yarasa və kələzin CoV viruslarının E, M və N struktur zülalları müqayisə olunmuşdur. İlk dəfə bu tədqiqatlarda aşkar olunan, ən təəccüblü fakt SARS-CoV-2 və kələz CoV-unun M və N zülallarındakı sinonim əvəzləmələrin nisbi payının digər koronaviruslarala müqayisədə əhəmiyyətli dərəcədə yüksək olmasıdır. Bu fakt SARS-CoV-2-nin yaranması və divergensiya dövrü ilə bağlı yeni suallar doğurur.

Açar sözlər: *Koronavirus, COVID-19, E zülalı, M zülalı, N zülalı, kələz, yarasa, SARS-CoV-2-nin mənşəyi, sinonimik mutasiyalar, sinonimik olmayan mutasiyalar*

Сравнительные исследования структурных белков E, M и N SARS-CoV, SARS-CoV-2, CoV панголина и летучей мыши

К.Г. Гасымов[1], Т.А. Самадова[1], Ф.К. Абасова[2], И.А. Шахмурадов [1,3*]

[1] Институт биофизики НАН Азербайджана, Баку, Азербайджан
[2]Лицей имени академика Зарифы Алиевой, Баку, Азербайджан
[3] Институт молекулярной биологии и биотехнологий НАН Азербайджана, Баку, Азербайджан

В течение последних двух десятилетий человечество страдает от 3 коронавирусных заболеваний, хотя коронавирусы человека были обнаружены более 50 лет назад. Последнее коронавирусное заболевание, обнаруженное в 2019 году (COVID-19), вызвано коронавирусом 2 тяжелого острого респираторного синдрома человека (SARS-CoV-2). Возникает вопрос: в чем может быть причина такой активации коронавирусов в последние годы? Чтобы ответить на этот вопрос, по крайней мере, необходимо выяснить (1) историю и происхождение этих вирусов, (2) изучить молекулярные механизмы их очень легкого и быстрого проникновения в клетки-хозяева, влекущего за собой многогранные серьезные нарушения. В этом исследовании мы сравнили структурные белки E, M и N из вирусов SARS-CoV-2, SARS, CoV летучих мышей и панголинов. Наиболее поразительный факт, впервые обнаруженный в этом исследовании, заключается в том, что относительная доля синонимичных скоростей замен в M и N белках SARS-CoV-2, и CoV панголина значительно выше, чем соответствующие характеристики для других изученных CoV. Это факт ставит несколько интригующих вопросов о возникновении и продолжительности дивергенции SARS-CoV-2.

Ключевые слова: *Коронавирус, COVID-19, E-белок, M-белок, N-белок, ящер, летучая мышь, происхождение SARS-CoV-2, синонимичные мутации, несинонимичные мутации*