In silico comparative genomics of SARS-CoV-2 to determine the source and diversity of the pathogen in Bangladesh

Tushar Ahmed Shishir [1¶], Iftekhar Bin Naser [1&*] Shah M Faruque [2&]

[1] Department of Mathematics and Natural Sciences, BRAC University, Dhaka, Bangladesh.
[2] School of Life Sciences, Independent University Bangladesh, Dhaka, Bangladesh.

* Corresponding author

Email: iftekhar.naser@bracu.ac.bd

[¶] This author outlined and carried out the analysis, prepared and edited manuscript

[&]These author revised, edited and approved the manuscript

Abstract:

The COVID19 pandemic has originated in China, immobilized the whole world including Bangladesh. Herein using whole-genome sequencing data of 64 COVID 19 viruses from Bangladesh, we tried to predict the possible viral transmission route to Bangladesh and divergence among theses sequences. Phylogenetic analysis indicated that the pathogen arrived in Bangladesh from multiple countries. The virus of Chattogram was found to be originated from Saudi Arabia but Dhaka was infected with the viruses from the United Kingdom and France. The total 64 sequences of Bangladesh generate three clusters. Besides, 180 mutations were present in the coding regions of these viruses, and 110 of them were missense. Among the missense mutations 99 were predicted to destabilize the protein structures. Specifically a mutation at NSP2 1163A>T (I300F) was highly prevalent in Bangladeshi sequences and predicted to have a destabilizing effect on NSP2, which modulates host cell survival strategy.

Keywords: SARS-CoV-2, COVID19, Bangladesh, Genome, Coronavirus, SARS-CoV-2 Variants, SARS-CoV-2 Natural Selection

Introduction.

The pandemic of coronavirus disease referred to as COVID-19 pandemic, which originated in China in December 2019 is ongoing and has now spread to 213 countries and territories. As of July 2020, the pandemic has caused about 10 million cases and half a million death. This novel virus of the Coronaviridae family and Betacoronavirus genus (1, 2) designated as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is the causative agent of COVID-19. Previously two other coronaviruses SARS-CoV and MERS-CoV have demonstrated high pathogenicity and caused epidemic with mortality rate ~10% and ~34% respectively affecting more than 25 countries each time (3–5). However, SARS-CoV-2 has proved to be highly infectious and caused pandemic spread to over 213 countries and territories. Besides its devastating impact in North America and Europe, the disease is now rapidly spreading in South America including Brazil, and in South Asian countries particularly India, Pakistan and Bangladesh (6).

The virus was first detected in Bangladesh in March 2020 (7). Although infections remained low until the end of March it began to rise steeply in April. By the end of June, new cases in Bangladesh grew to nearly 150,000 and the rate of detection of cases compared to the total number of samples tested increased to about 22% which was highest in Asia (6).

SARS-CoV-2 is a positive-sense single-stranded RNA virus with a genome size of nearly 30kb. The 5' end of the genome codes for a polyprotein which is further cleaved to viral non-structural proteins whereas the 3' end encodes for structural proteins of the virus including surface glycoproteins spike (S), membrane protein (M), envelop protein (E) and nucleocapsid protein (N) (8). Like other RNA viruses, SARS-CoV-2 is also inherently prone to mutations due to high recombination frequency resulting in genomic diversity (9–11). Due to the rapid evolution of the

virus, development of vaccines and therapies may be challenging. To monitor the emergence of diversity, it is important to conduct comparative genomics of viruses isolated over time and in various geographical locations. Comparative analysis of genome sequences of various isolates of SARS-CoV-2 would allow to identify and characterize the variable and conserved regions of the genome and this knowledge may be useful for developing effective vaccines. Thousands of SARS-CoV-2 virus genome have been sequenced and submitted in public databases for further study. This include 66 SARS-CoV-2 genomic sequences submitted from Bangladesh in the Global initiative on sharing all influenza data (GISAID) database, till 11th June 2020 (12). We conducted comparative analysis of publicly available genome sequences of SARS-CoV-2 from 27 countries to predict the origin of viruses in Bangladesh by studying a time-resolved phylogenetic relationship. Later, we analyzed the variants present in different isolates of Bangladesh to understand the pattern of mutations, find unique mutations, and possible effect of these mutations on the stability of encoded proteins, and selection pressure on genes.

Materials and methods

A total of 435 whole genome sequence of SARS-CoV-2 including 64 sequences isolated in Bangladesh (detail information in S1 Table), and that of 5 isolates of each month between January to May, 2020 selected from 27 different countries with high number of infection were included in this analysis. Source and sequence number are presented in Table 1 (detail information are provided in S2 table). However, African countries had limited number of sequences therefore all the sequences from African countries were categorized as African sequence (12). Reference sequence included the first sequence from Wuhan (NC_045512.2) (13). Selected sequences were annotated by Viral annotation pipeline and Identification tool (VAPiD)

(14), and multiple sequence alignment was carried out using Mafft algorithm (15). Maximum likelihood phylogenetic tree was constructed with IQ-TREE (16), the generated tree was reconstructed based on time-calibration by TreeTime (17), and visualized on iTOL server (18). For mutation analysis, sequence were mapped with minimap2 (19), and variants were detected using SAMtools (20) and snp-sites (21). A haplotype network was generated based on mutations in genome using PopArt (22). Sequences were then put into different clades based on specific mutations proposed in GISAID (23) and further classified as D614G type (24, 25). Then another phylogenetic tree and haplotype network containing only Bangladeshi sequence was constructed and categorized using the same pipeline, additionally one step further clustered with TreeCluster (26). The direction of selection in Bangladeshi sequences was calculated by the SLAC algorithm (27) in the Datamonkey server (28). Last, of all, the effects of the mutations on protein stability were predicted using DeepDDG (29).

Table 1.Sequences taken from different countries and regions of the World

Country	Number of Sequences
1. African region	17
2. Australia	27
3. Bangladesh	65
4. Brazil	15
5. Canada	19
6. China	31
7. France	20
8. Germany	20
9. India	18

10. Iran	5
11. Italy	30
12. Mexico	8
13. Nepal	1
14. Pakistan	3
15. Russia	14
16. Saudi Arabia	19
17. Spain	15
18. Sri Lanka	6
19. Sweden	5
20. Turkey	10
21. United Kingdom	31
22. USA	55

Results and Discussion

Phylogenetic analysis of all SARS-CoV-2 sequences

Genomic Sequences of 64 SARS-CoV-2 isolates from Bangladesh were analyzed in the time-resolved phylogenetic tree containing reference (NC_045512.2 first Chinese SARS-CoV-2 sequence) and 433 sequences reported from various countries of the world. Sequences from Bangladesh formed three clusters of which one cluster carried 43 of the 64 sequences, and shared the same node with sequence from Germany and they had a common ancestor with sequences from the United Kingdom (Fig 1). Two other clusters of Bangladesh had 4 and 5 sequences respectively, and in both cases, they shared the same node with sequence of SARS-CoV-2 reported from India and further centered and shared a common ancestry with those from Saudi Arabia. On the other hand, 12 alone sequences that did not belong to any clusters were found to be similar with sequences from Europe including United Kingdom, Germany, France, Italy, and Russia while one of these sequences was close to sequence reported from the USA.

Subsequently, all SARS-CoV-2 sequences from representative countries were clustered based on some specific mutations sustained into different clades as mentioned by GISAID. In this analysis, the sequences from Bangladesh were found to be distributed in all clades except V (Figs 1 and 2).

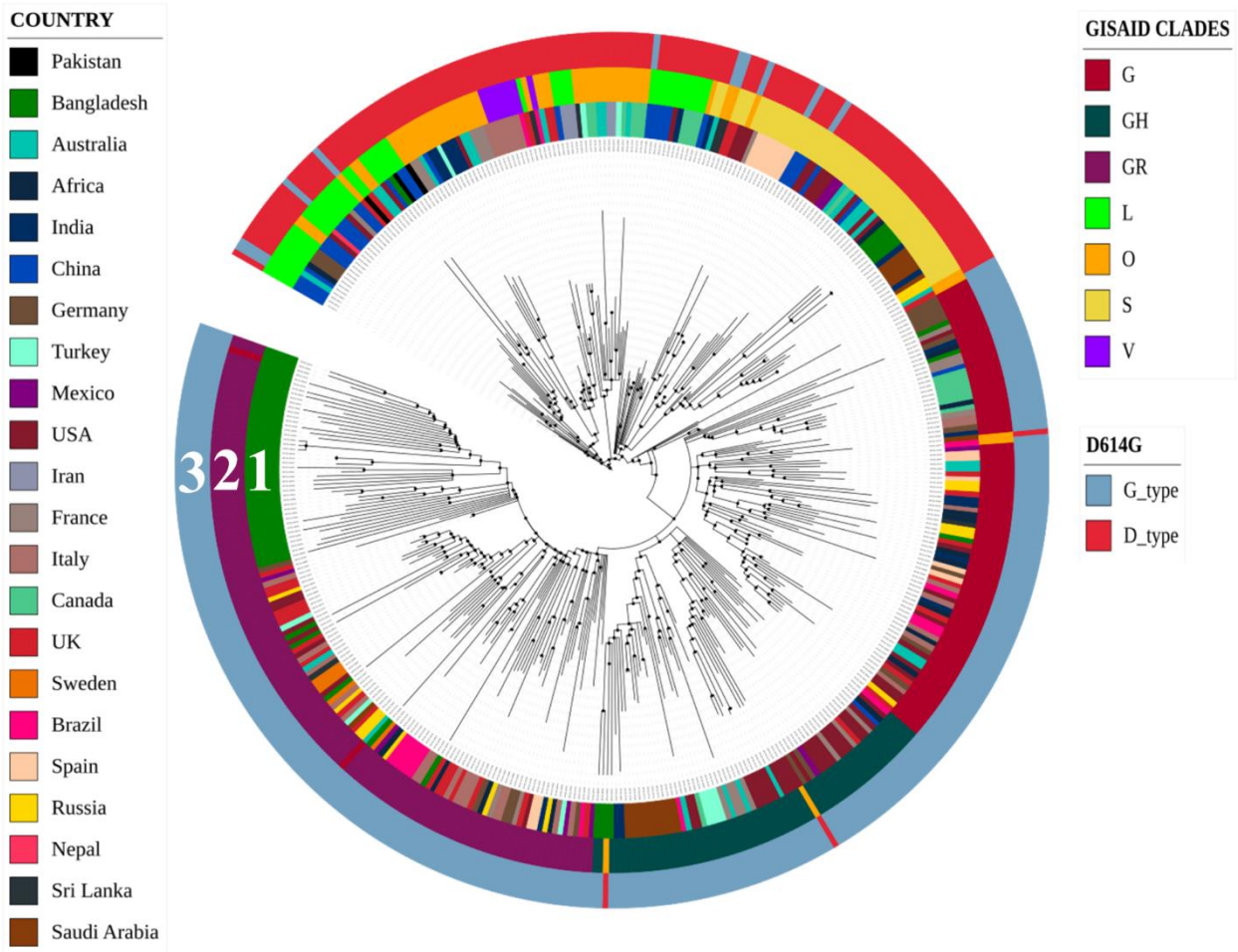

Fig 1. **The time-resolved phylogenetic tree, containing sequences from different countries.** Circle 1 exhibits the location of the sequences and how these are distributed; circle 2 demonstrates GISAID clades based on some specific mutations and circle 3 represents D614G classification.

Phylogenetic comparison of SARS-CoV-2 sequences from Bangladesh

In order to understand the evolutionary relationship and possible transmission dynamics of the isolates in Bangladesh, another time-resolved phylogenetic tree carrying only sequences of SARS-CoV-2 isolated in various districts of Bangladesh with reference (first SARS-CoV-2 sequence from China) was generated. This analysis produced three clusters in which cluster-1 had sequence of some isolates from Chattogram and that of one from Dhaka, cluster-3 had sequences only from Chattogram and cluster-2 had sequence from Dhaka, Narayanganj and Chattogram. The sequences from Bangladesh were found to be distributed in all GISAID clades except V. Most of the sequences of Dhaka andNarayanganj were from the GR clades (47 out of 52), sequences of SARS-CoV-2 from Chattogram belonged to five different clades (G, GH, GR, O, and S).

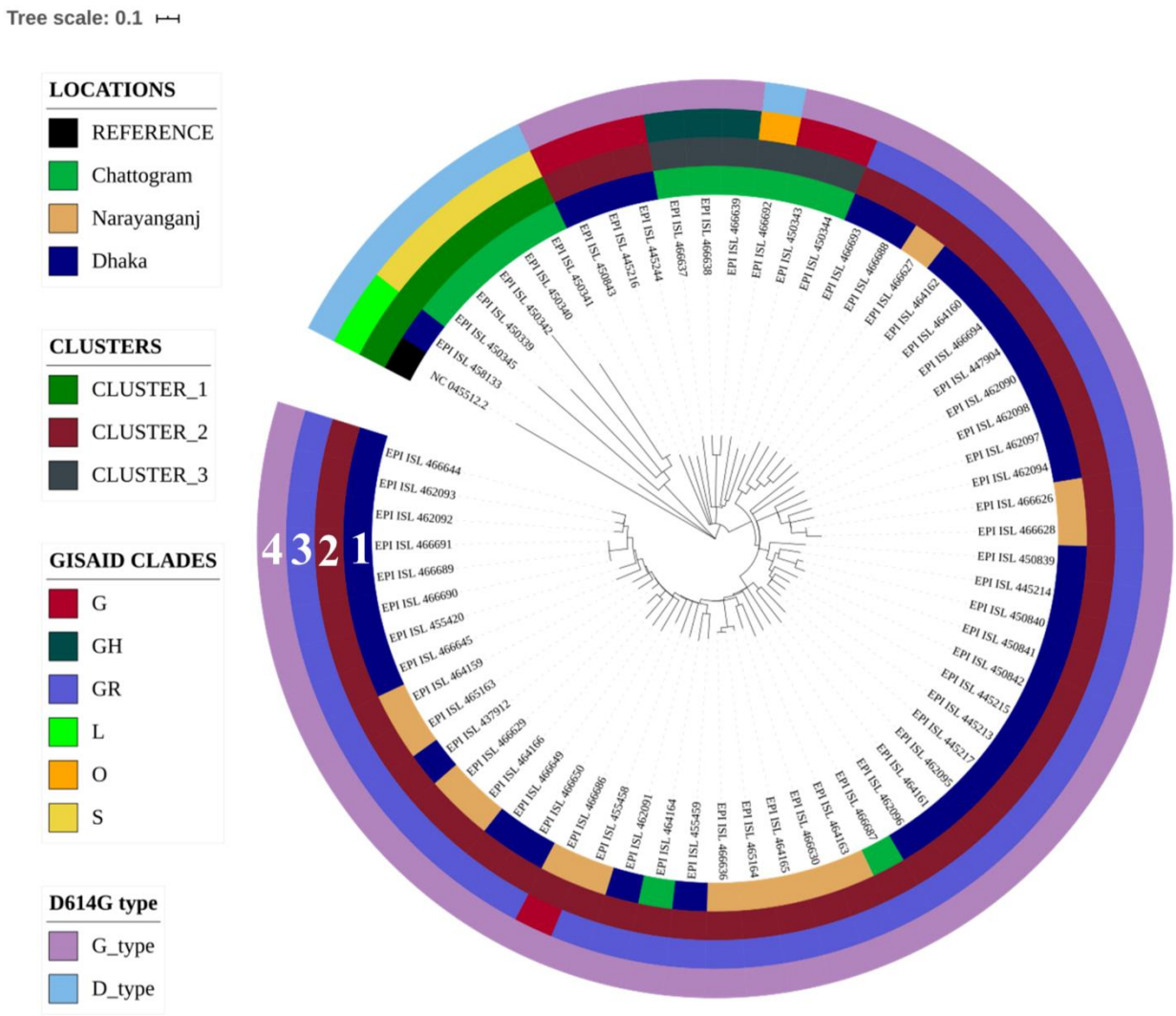

Fig 2. Time-resolved phylogenetic tree of Bangladeshi sequences.

Circle 1 represents the location of sample isolation, circle 2 divides the sequences of the three clusters based on their intimacy, circle 3 distributes in sequences in GISAID clades and circle 4 represents if the sequences are D type or G type

D614G mutations in spike protein gene

Since 23403A>G mutation in the spike (S) protein has been suggested to have correlation with increased transmission, the SARS-CoV-2 sequences were also categorized according to D614G type mutation (Fig 1). This particular subtype with a non-silent (Aspartate to Glycine) mutation at 614th position of the Spike protein rapidly outcompeted other preexisting subtypes, including the ancestral strain. The D614G mutation generates an additional serine protease (Elastase) cleavage site near the S1-S2 junction of the Spike protein (25). All but one sequence from Dhaka and Narayanganj were found to be of G type which carries Glycine at position 614 whereas sequences of Chattogram carry sequences of both types (Fig 2). In addition, the first sequence of Bangladesh carried G614 type of surface glycoprotein, indicate this dominant variant was present from the beginning may be due to the mutant virus was carried to the country from Europe which made the viral transmission easy.

Therefore, considering the phylogenetic analysis, the results demonstrate that all the isolates of Dhaka are the descendant of European countries, more specifically France and the United Kingdom. On the other hand, isolates of Chattogram are least distant from Saudi Arabian isolates. Moreover, the presence of S clade is absent among Dhaka isolates but most of the strains of Chattogram belong to S clade, which demonstrates that the sources of infection in these two areas are different. Then, isolates of Narayanganj and two isolates of Dhaka have a close association, indicates Dhaka is acting as the hub of infection inside the country.

Haplotype analysis of SARS-CoV-2 sequences

Relationships among DNA sequences within a population is often studied by constructing and visualizing a haplotype network. We constructed a haplotype network by the Median Joining algorithm and found that 338 of 434 SARS-CoV-2 sequences from representative countries were alike, therefore formed a large haplo group (Fig 3A). However, there were presence of a significant number of unique lineages too consisting of a single or multiple SARS-CoV-2 sequences (Fig 3A). This network demonstrated the closeness of the sequences and their pattern of mutation beyond the geographical boundary. Several SARS-CoV-2 isolates appeared to have sustained certain common mutations along with certain unique mutations. Although a large proportion of sequences from Bangladesh belonged to the common cluster (Fig 3A), there was a significant number of unique nodes as well due to mutations overtime after being carried into Bangladesh (Fig 3A). Therefore analysis of the sequences from Bangladesh provided further insight of their mutations and the haplotype network revealed that viruses isolated in Bangladesh had unique mutations in them, and as a result they belonged to different haplo groups and no significant cig group (Fig. 3A). Most of the sequences sustained a significant number of mutation compared with each other. Furthermore, it appears most isolates from Chattogram (Fig 3B) were not directly related to those isolated in Dhaka or Narayanganj.

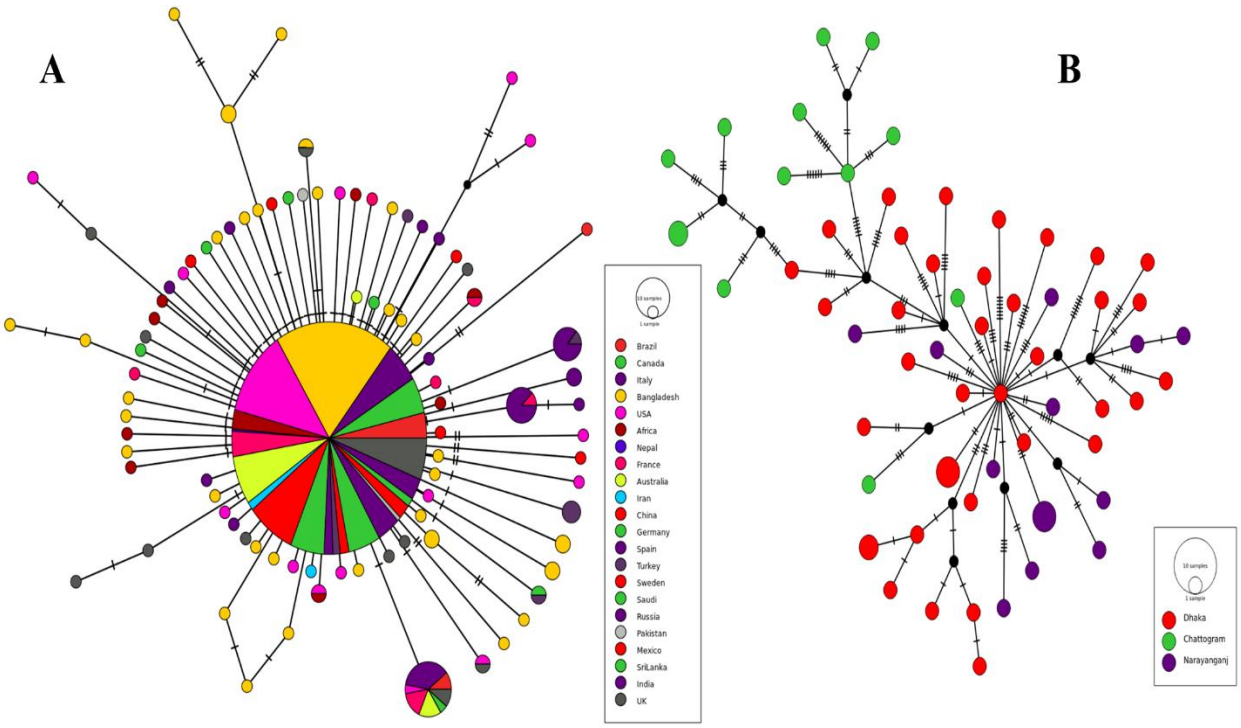

Fig 3. Haplotype network of selected sequences representing relatedness considering mutations.

(A) The biggest circle represents a large group of sequences from different countries that are similar and few sequences are connected with other sequences through undetermined intermediate due to harboring unique mutations denoted by small black circle. (B) Haplotype network of Bangladeshi sequences, the number of dashes in the connecting line denotes the number of mutations against each other.

Location and predicted effects of the mutations.

We detected the presence of 209 point mutations in 64 SARS-CoV-2 genome sequences reported from Bangladesh when compared to the reference sequence from China. Then, 19 sequences were found to lose significant portions of their genome as a result some viruses missed non-structural proteins such as ORF7 and ORF8 while other deletions were upstream or downstream

gene variants (S3 Table). Among the point mutations, 29 mutations were in the non-coding region of the genome and 180 were in coding regions. Ten of the 29 non-coding mutations were in upstream non-coding region and rest was in downstream non-coding region of the genome. Seventy mutations in the coding region were synonymous and 110 mutations predicted substituted amino acids. Among twelve predicted ORFs, ORF1ab which comprises approximately 67% of the genome encoding 16 nonstructural proteins had more than 60 percent of the total mutations while gene E encoding envelope protein and ORF7b were conserved and did not carry any mutation. Though ORF1 harbored the highest number of mutations but mutation density was highest in ORF10 considering gene length. Details and distribution of the mutations are presented in Table 2 and full analysis report is placed in S3 table.

Table 2. Count of mutation in each gene, types of mutation, mutation density and directional selection value. (dN/dS> 1 is positive selection, dN/dS = 1 neutral selection, dN/dS< 1 negative selection, dN/dS = 0 is conserved region)

ORF	**ORF length**	**Total No. of mutation**	**No. of Synonymous Mutations**	**No. of Missense Mutations**	**Mutation density**	**dN/dS**
ORF1ab	21291	110	44	66	0.52%	0.49
ORF1a	13219	84	32	52	0.64%	0.63
S	3823	17	5	12	0.44%	0.79
ORF3a	829	13	4	9	1.57%	0.89
E	229	0	0	0	0.00%	NA

M	670	7	5	2	1.04%	0.23
ORF6	187	3	0	6	1.60%	17.8
ORF7a	367	1	1	0	0.27%	1.08
ORF7b	133	0	0	0	0.00%	NA
ORF8	367	9	5	4	2.45%	0.57
N	1261	17	3	14	1.35%	1.55
ORF10	118	3	3	0	2.54%	0

The 241C>T and 3037C>T were the two most abundant mutations found to in 58 out of 64, and often found simultaneously in sequences from Bangladesh. Position 241 position is in the non-coding region whereas the mutation in position 3037 was synonymous. On the other hand, 57 sequences were found to harbor 14408C>T and 23403A>G mutations which altered amino acid Pro>Leu and Asp>Gly respectively, and these two mutations were found to be present simultaneously as well. In addition, other co-evolving mutations found were (241C>T, 3037C>T, 14408C>T, 23403A>G), (28881G>A, 28882G>A, 28883G>C), (8782C>T, 28144T>C), (4444G>T, 8371G>T, 29403A>G). Apart from highly abundant mutations, several less common and unique mutations were also present in the sequences analyzed.

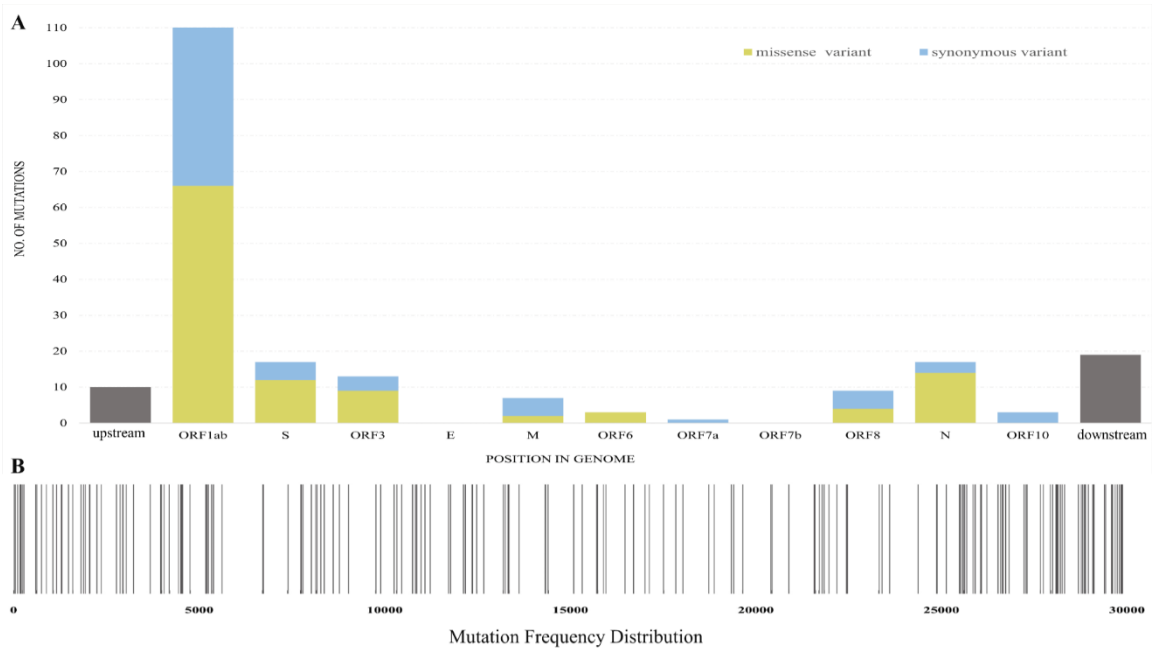

Fig 4. Mutations present in SARS-COV-2 genome sequences reported from Bangladesh

(A) Number of mutations at different regions (B) Mutation distribution frequency

Then again, as a result of continuous mutations genes were going through natural selection pressure. ORF6 predicted was to have dN/dS value 17.8 due to the presence higher number of missense than synonymous mutations indicate that, ORF6 is rapidly evolving under high selection pressure and highly divergent. Then again, ORF7a and Nucleocapsid phosphoprotein had dN/dS values 1.08 and 1.55 respectively which confer their strong evolution to cope up with challenges under positive selection pressure. ORF10 is predicted to be conversed with dN/dS value 0 while envelope protein and ORF7b did not harbor any mutation due to conservancy. On the other hand, ORF3a and surface glycoprotein might approach toward positive selection pressure and evolve but rest other proteins were under negative selection pressure.

Table 3. High-frequency mutations present in Bangladeshi sequences and their effect on protein's structure stability. (ddG< 0: decrease stability, ddG> 0: increase stability)

Position	Gene	Feature	Amino acid change	Frequency	ddG (kcal/mol)
241 C > T	5' UTR	Non-coding	-	58/64	
1163 A > T	ORF1ab	missense	I300F	44/64	-0.756
3037 C > T	ORF1ab	Synonymous	-	58/64	
14408 C > T	ORF1ab	Missense	P4715L	57/64	1.463
23403 A > G	S	Missense	D614G	57/64	-0.45
28881 G > A	N	Missense	R203L	49/64	-3.4
28882 G > A	N	Synonymous	-	49/64	
28883 G > C	N	Missense	G204A	49/64	-4.3

Among 110 missense mutations, 99 predicted to destabilize the proteins while only 11 mutations increase stability. Stabilizing 10 Mutations were present in ORF1 operon and 1 mutation at ORF3a. On the other hand, destabilizing mutations were distributed among most of the ORFs. None of the mutations in structural proteins were predicted to increase stability.

Table 4. Mutations present in different ORFs present in Bangladeshi sequences

ORF	Unique Mutation for Bangladesh	Remarks
ORF1ab	93	A mutation at 1163A>T (Ile300Phe) in ORF1 operon which lower the structural stability of the protein was found existing in 44

		strains of Bangladesh but absent in representative countries
S	13	A highly prevalent mutation 23403A>G (Asp614Gly) which ease the viral transmission, dominating throughout the world was found to be present in 57 of 64 sequences in Bangladesh.
ORF3a	9	A mutation at 25563G>T (Gln57His) was found highly prevalent among selected sequences from the USA was also present in Bangladeshi sequences.
M	4	Unique mutations were synonymous
ORF6	3	3 missense mutations were absent in sequences from representative countries.
ORF7a	1	Synonymous mutation
ORF8	7	3 missense mutations but their frequency was lower
N	10	Mutations at 28881G>A (Arg203Lys), 28882G>A (Arg203ARG), 28883G>C (Gly204Arg) were harbored in 49 of 64 sequences which were found highly prevalent among viruses in European and North American countries
ORF10	3	All mutations were synonymous and found this gene conserved

Therefore, considering the outcome of mutations analysis we have found point mutations as well as basepairs deletion. Deletions of the basepairs were associated with missing non-structural proteins and affected the viral properties since ORF7a is the growth factor of the coronavirus family, induce apoptosis, and promotes viral encapsulation (30–32) while ORF8 is associated

with viral adaptation by playing role in host-virus interaction (33,34). Furthermore, we have found that some genes are under positive selection pressure and evolving rapidly while some other genes are progressing toward a positive selection indicating the virus is fast-evolving to evade host cell's innate immunity; which should be taken into special consideration prior to vaccine development or other treatment strategies.

Lastly, a missense mutation at 1163A>T changing the amino acid isoleucine to phenylalanine of nsp2 found unique among 44 Bangladeshi strains. Nsp2 is a methyltransferase like domain that interacts with PHB and PHB2, modulates the host cell survival strategy by affecting cellular differentiation, mitochondrial biogenesis, and cell cycle progression to escape from innate immunity (34,35). To recapitulate, this unique and high-frequency mutation might be a further interest of study considering death rate against the infection rate in Bangladesh.

References.

1. WHO. Coronavirus disease COVID-2019 - Situation Report 169. World Heal Organ. 2020;

2. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res. 2018;

3. Zhong NS, Zheng BJ, Li YM, Poon LLM, Xie ZH, Chan KH, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet. 2003;

4. De Wit E, Van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: Recent insights into emerging coronaviruses. Nature Reviews Microbiology. 2016.

5. Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. Respirology. 2018.

6. Webmeter. Coronavirus Age, Sex, Demographics (COVID-19) - Worldometer. www.worldometers.info. 2020.

7. Rahaman Khan MH, Hossain A. COVID-19 Outbreak Situations in Bangladesh: An Empirical Analysis. medRxiv. 2020.

8. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nature Reviews Microbiology. 2019.

9. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. Coronaviruses: An RNA proofreading machine regulates replication fidelity and diversity. RNA Biology. 2011.

10. Yang Y, Liu C, Du L, Jiang S, Shi Z, Baric RS, et al. Two Mutations Were Critical for Bat-to-Human Transmission of Middle East Respiratory Syndrome Coronavirus. J Virol. 2015;

11. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nature Medicine. 2020.

12. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance. 2017.

13. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;

14. Shean RC, Makhsous N, Stoddard GD, Lin MJ, Greninger AL. VAPiD: A lightweight cross-platform viral annotation pipeline and identification tool to facilitate virus genome submissions to NCBI GenBank. BMC Bioinformatics. 2019;

15. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. MolBiolEvol. 2013;

16. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. MolBiolEvol. 2015;

17. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 2018;

18. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;

19. Li H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 2018;

20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence

Alignment/Map format and SAMtools. Bioinformatics. 2009;

21. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb genomics. 2016;

22. Leigh JW, Bryant D. POPART: Full-feature software for haplotype network construction. Methods EcolEvol. 2015;

23. GISAID. Clade and lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses [Internet]. 2020 [cited 2020 Jun 21]. Available from: https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hCoV-19-viruses/

24. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020;

25. Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv. 2020;

26. Balaban M, Moshiri N, Mai U, Jia X, Mirarab S. TreeCluster: Clustering biological sequences using phylogenetic trees. PLoS One. 2019;

27. Kosakovsky Pond SL, Frost SDW. Not so different after all: A comparison of methods for detecting amino acid sites under selection. MolBiolEvol. 2005;

28. Kosakovsky Pond SL, Frost SDW. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics. 2005;

29. Cao H, Wang J, He L, Qi Y, Zhang JZ. DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks. J ChemInf Model. 2019;

30. Hänel K, Willbold D. SARS-CoV accessory protein 7a directly interacts with human LFA-1. Biol Chem. 2007;

31. Schaecher SR, Touchette E, Schriewer J, Buller RM, Pekosz A. Severe Acute Respiratory Syndrome Coronavirus Gene 7 Products Contribute to Virus-Induced Apoptosis. J Virol. 2007;

32. Taylor JK, Coleman CM, Postel S, Sisk JM, Bernbaum JG, Venkataraman T, et al. Severe Acute Respiratory Syndrome Coronavirus ORF7a Inhibits Bone Marrow Stromal Antigen 2 Virion Tethering through a Novel Mechanism of Glycosylation Interference. J Virol. 2015;

33. Zhang Y, Zhang J, Chen Y, Luo B, Yuan Y, Huang F, et al. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. bioRxiv. 2020;

34. Yoshimoto FK. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. Protein J [Internet]. 2020;39(3):198–216. Available from: https://doi.org/10.1007/s10930-020-09901-4

35. Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. J Med Virol. 2020;

Supporting information

S1 Table. Sequences from Bangladesh.

S2 Table. Sequences from selected countries.

S3 Table. Mutations in Bangladeshi sequences.