Bangladesh Pharmaceutical Journal ***26****(2): 204-211, 2023 (July)*

Mutational and Phylogenetic Analyses of SARS-Cov-2: Bangladesh Perspective

B. M. Fahim Faisal[1] and Md. Selim Reza[2]

[1]Deptartment of Pharmacy, University of Dhaka, Dhaka-1000, Bangladesh
[2]Department of Pharmaceutical Technology, Faculty of Pharmacy, University of Dhaka
Dhaka-1000, Bangladesh

(Received: May 15, 2023; Accepted: July 20, 2023; Published (web): July 25, 2023)

Abstract

Severe acute respiratory syndrome or SARS-Cov-2 is a global pandemic causing over 6.5 million deaths worldwide. As a part of the global village, Bangladesh is also heavily affected with more than 29,000 deaths in those two years of pandemic. Though the severity and mortality rate are not so high compared to the other countries, Bangladesh also faced many difficulties combating the mutated and newer types of the Covid-19. The evolution and phylogenetic study of the virus is one of those key areas to focus in order to understand the nature of virus, its mutation pattern and gene prediction. This study investigates the evolutionary links among the older and newly emerged SARS-Cov-2 variants. In this study, the entire genome sequences of SARS-Cov-2 variants were obtained, aligned using muscle alignment, pairwise comparison was computed, differences, gaps and mutations were noted. The phylogenetics of different types of covid-19 variants were determined using a variety of evolutionary substitution models. The ultrametric and metric clustering methods, such as UPGMA and neighbor-joining (NJ), using nucleotide substitution models allowed the inclusion of nucleotide transitions and transversions as Kimura 80 models. The findings revealed that Omicron variant forms a new monophyletic clade that is distant from other SARS-Cov-2 variants but surprisingly in proximity to the alpha variant, which was actually dominant in the early stage of pandemic. This finding also indicates that the omicron variant might have been there for a while, hiding the virulence and later becomes the variant of concern in the latest wave of the pandemic. This may pave the way for new researchers to find proper insight of different variants of Bangladesh in treatment plan and vaccine designing.

Key words: Covid-19, genome evolution, phylogenetics, nucleotide sequence, mutation.

Introduction

Severe acute respiratory syndrome or SARS-Cov-2 was first found in Wuhan, China in 2019 as a result it is also known as the Covid-19. The viruses of the Coronaviridae subfamily are divided into four genera, one of which is the beta-coronavirus (Erles *et al.,* 2003). There are four different lineages of the beta-coronavirus: 2a, 2b, 2c, and 2d. Members of the 2b lineage include SARS-CoV and SARS-Cov-2, while the 2c lineage includes MERS-Cov (Liya *et al.,* 2020). It is not yet known how SARS-Cov, MERS-Cov and SARS Cov-2 spread from animals to humans. Bats are believed to be the natural source of virus transformation, also known as bat-to-man SARS-Cov-2 transformation, because 96.2 percent of the SARS Cov-2 gene is similar to the bat genome. Ordinarily, SARS-CoV is constructed similarly to other coronaviruses (Zhou *et al.,* 2020).

The virus binds to the human ACE-2 receptor causing fever (81.2%), cough (58.5%), fatigue (38.5%), dyspnea (26.1%), and the sputum (25.8%.) (Alimohamadi *et al.,* 2020) The SARS-Cov-2 has an

Corresponding author: Md. Selim Reza; Email: selimreza@du.ac.bd; Phone: +88 01715 363765
DOI: https://doi.org/10.3329/bpj.v26i2.67810

affinity towards ACE-2 receptors 10 times greater than that of SARS-Cov (Yuan *et al.*, 2017), as a result the spreadability is much more in case of Covid-19. The first recorded news of covid-19 in Bangladesh was in March 8, 2020. The main reported cause for the outbreak of the virus in Bangladesh is the immigrant people of Bangladesh who came from different countries, especially from Italy, UK and China. Also, the density of population and overpopulation exacerbated the situation, since it is a highly contagious virus. The country also faced difficulties with the evolution and emergence of the new variants with shortage of ICUs, oxygen supply, scarcity of testing kits and so on. Up to July 2022, there are 1.98M covid cases with 29,185 deaths.

It was delta variant that is responsible for the worst scenario of death in Bangladesh in the mid 2021 as shown in figure 1 below. Again, there were issues with the vaccine supply and its management and the country had to look for different sources of vaccine from different parts of the world.

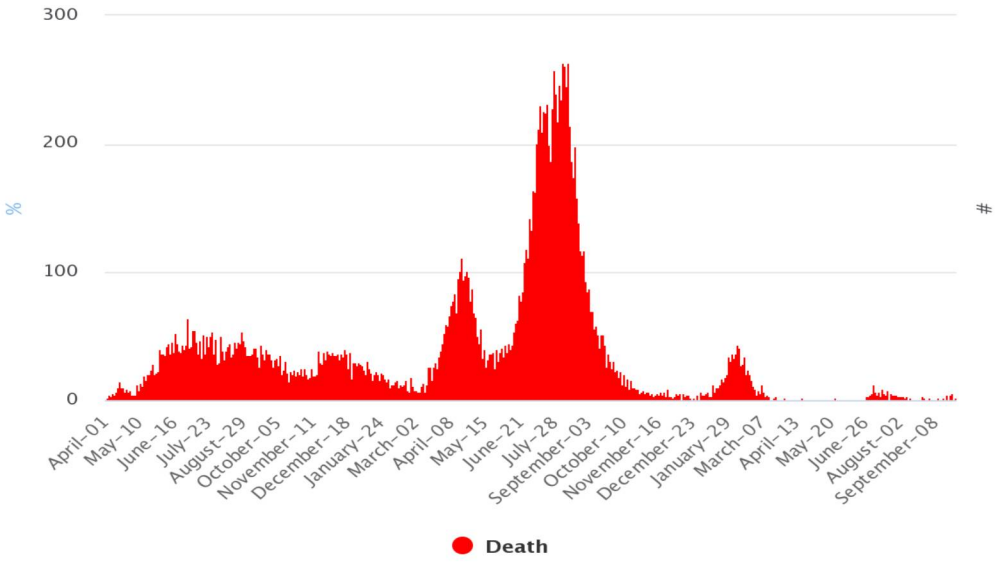

Figure 1. Covid-19 deaths in Bangladesh.

In this study, an attempt was taken to observe the phylogeny of these variants of concerns and the prevalence of those variants with the timeline of covid-19 in Bangladesh. Different methodologies were for this purpose.

Materials and Methods

Genome collection and analytical program: Preliminary data was analyzed from the Global Initiative on Sharing Avian Influenza Data, GISAID (http://www.gisaid.org/). Upon selection to Bangladesh with complete sequence, 6,878 submissions were found. Of them, alpha (103), beta (490), Delta (3098), Omicron (1593) variants consist almost 99% of the data. Among these variants, B.1.1.25, B.1.1, B.1.617.2, B.1.1.7, AY.122, BA.1 and BA.2 were found to dominant as shown in the figure 2.

Nucleotide sequences of these variants were derived from NCBI virus (https://www.ncbi.nlm.nih.gov/). The geographic region was selected to Bangladesh and only complete genome sequences were selected.

In 2020, B.1.1.25 (alpha variant), also known as the Bangladesh variant (Web-1) was dominant, followed by B.1.1 and B.1. On the other hand, B.1.617.2 (Delta variant) was dominant, followed by AY.122 (Alias of B.1.617.2.122, European lineage) (Web-1) and B.1.1.7 in the year 2021. BA.2 (omicron variant) was found in the late September. But

Omicron was dominant in the year 2022. Among the omicron variants, BA.1 and BA.2 were found to be dominating. So, the nucleotide sequence (FASTA format) of the following dominant variants, along with the reference Wuhan were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) in order to obtain phylogenetic relationship between them.

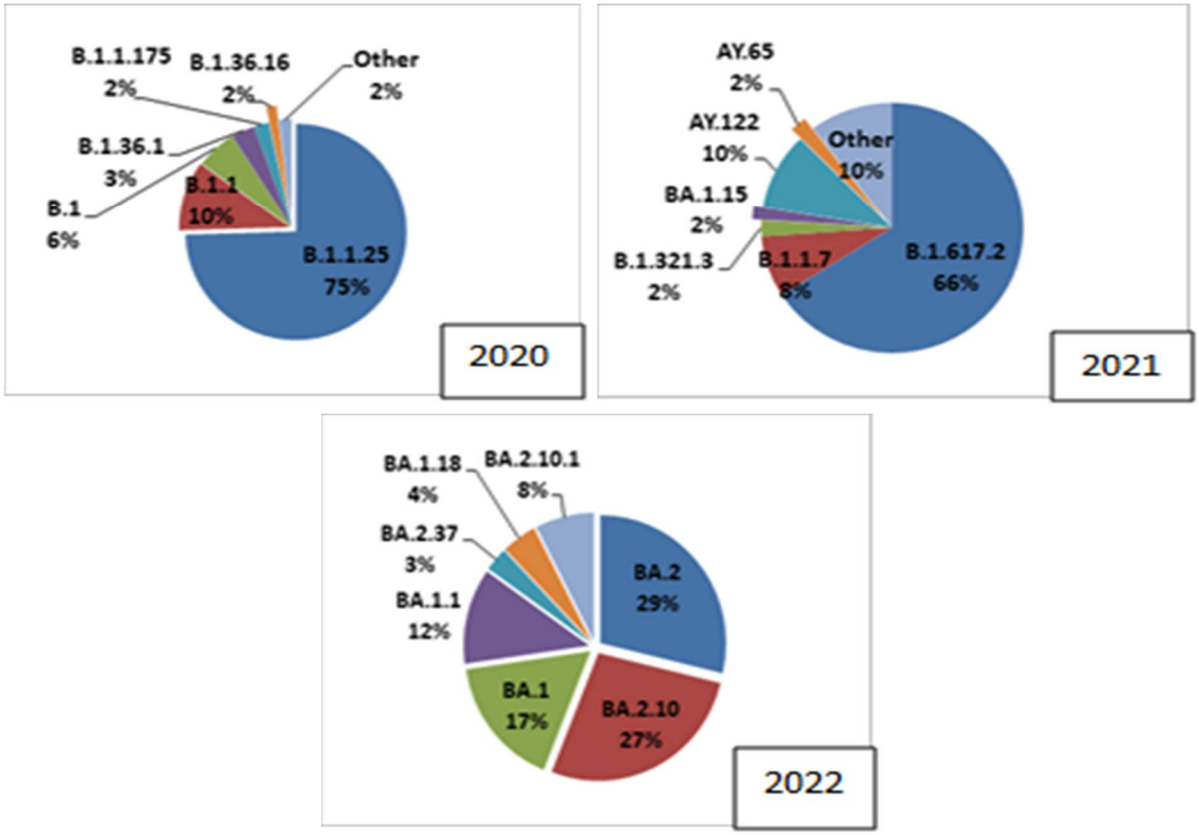

Figure 2. Yearly prevalence of different Covid-19 variants. From the left, year 2020, 2021 and 2022.

Table 1. List of Covid-19 variants used in this study including pango lineage, accession ID in NCBI virus, their collection and release date and length.

Variant	Accession ID	Collection date	Release date	Length	Host
B.1.1.25	MT775558	2020-06-29	2021-01-21	29834	Human
B.1.1	MZ749908	2020-06-11	2021-08-11	29873	Human
B.1.617.2	OM049448	2021-08-01	2021-12-29	29811	Human
B.1.1.7	MZ413975	2020-12-31	2021-07-01	29766	Human
AY.122	OL911013	2021-06-29	2021-12-19	29814	Human
BA.2	ON026034	2022-01-20	2022-03-21	29760	Human
BA.1	ON026035	2022-01-20	2022-03-21	29776	Human

Selection criteria of the variants: The genome sequences were downloaded from NCBI upon selection of the up-to-date coronavirus submitted. The geographic region was selected to Bangladesh. Only the full-length entries were counted, and no sequence less than 29000 nucleotide was selected. The

reference genome sequence is to be the the one found in Wuhan back in 2019. They were sorted according to the pango lineages, the host to be human. Three different submission and collection times were taken because in this study, the evolution of coronavirus with time is studied. The selected sequences along with the accession ID, collection date, release date and size of the length is mentioned in table 1.

Alignment of nucleotide sequences: There are usually two main types of alignment. One is the local alignment used to gather the island of matches. It is more useful for dissimilar sequences that are suspected to contain regions of similarity or similar sequence motifs within their larger sequence context.

Another type of alignment is the global alignment. It is the attempt to the alignment of every residue in every sequence. It is most useful when the query set is similar of roughly equal size (This does not mean global alignments cannot end in gaps.) For example- The Needleman-Wunch alignment.

Sequence similarity and sequence identity are synonymous for nucleotide sequences. For protein sequences, however, the concepts are very different. Similarity of an alignment refers to the percentage of aligned residues that have similar physiochemical characteristics and can readily substituted for each other.

Alignment of multiple sequences or multiple sequence alignment (MSA) can be seen as a generalization of pairwise sequence alignment, instead of aligning two sequences.

In our study, a muscle alignment using Jukes-Cantor model was conducted. Jukes-Cantor model can be used where the nucleotides are uniformly distributed and independently substituted such that the probabilities for nucleotide substitutions are all the same and do not depend on the nucleotides. It assumes equal rate of mutation for transition and transversion.

The downloaded FASTA format nucleotide sequences were aligned using muscle alignment provided in the molecular evolutionary genetics analysis, MEGA-11 with default parameters and then analyzed. The results were recorded and saved for further analysis. Then pairwise differences were calculated using Jukes-Cantor model.

Spike protein mutation: The SARS-Cov-2 is a spherical shaped virion with a positive-stranded RNA viral genome of size 30 kb that is translated into structural and non-structural proteins. This spike protein is a trimer that is present on the surface of the coronavirus that plays a vital role in recognition of human host cell surface receptor angiotensin-converting enzyme-2 (ACE-2). This recognition is required for fusion of viral and host cellular membranes for transfer of the viral nucleocapsid into the host cells.

The FASTA sequences were uploaded to GISAID mutation app (https://gisaid.org/database-features/covsurver-mutations-app/), reference hCov-19/Wuhan/WIV04/2019 and the changes in the spike protein was recorded.

Results and Discussion

Pairwise distance calculation: All the sequences from table 1 were aligned with MEGA-11. Pairwise comparison matrix showed that the greatest number of mutations was recorded with the omicron variant that is currently predominant in Bangladesh. There were at least 8 mutants found of the omicron variant. At the same time, Omicron also showed the greatest number of changes in nucleotide.

For pairwise distance measurement, Jukes-Cantor model was used, and the following result was obtained.

The distance between these variants shown in tables 2 and 3 reflects that the difference from the originating Wuhan variant is the largest for the omicron variants (BA.2 and BA.1 respectively) followed by the delta and alpha variant. There are only 0.03-0.1% dissimilarity between the alpha and the Wuhan nucleotides, while the number is 0.12% and 0.215% for delta and omicron respectively. At the same time, omicron has a 0.286% variation from delta variants which was dominant just prior to the omicron, which suggests that omicron was not originated from delta, rather it was from a wild source (like mouse)[7] or other possible ways.

Table 2. Estimates of evolutionary divergence between sequences (Jukes-Cantor).

	Wuhan	AY.122	B.1.1.7	B.1.1.25	B.1.1	B.1.617.2	BA.1
Wuhan							
AY.122	0.00144						
B.1.1.7	0.00100	0.00212					
B.1.1.25	0.00030	0.00144	0.00084				
B.1.1	0.00026	0.00147	0.00097	0.00020			
B.1.617.2	0.00127	0.00070	0.00195	0.00127	0.00131		
BA.1	0.00185	0.00279	0.00222	0.00168	0.00181	0.00262	
BA.2	0.00215	0.00303	0.00253	0.00198	0.00212	0.00286	0.00158

Table 3. Estimates of evolutionary divergence between sequences (Maximum likelihood).

	Wuhan	AY.122	B.1.1.7	B.1.1.25	B.1.1	B.1.617.2	BA.1
Wuhan							
AY.122	0.00144						
B.1.1.7	0.00100	0.00212					
B.1.1.25	0.00030	0.00145	0.00083				
B.1.1	0.00026	0.00147	0.00097	0.00020			
B.1.617.2	0.00126	0.00070	0.00194	0.00127	0.00131		
BA.1	0.00185	0.00278	0.00222	0.00167	0.00181	0.00263	
BA.2	0.00216	0.00304	0.00253	0.00198	0.00212	0.00286	0.00153

Table 4. Total nucleotide composition of the variants.

Variant	T(U)	C	A	G
NC 045512.2 |Wuhan	32.08373742	18.36605023	29.94348393	19.60672842
OL911013.1 |AY.122	32.20299188	18.33702288	29.84839337	19.61159187
MZ413975.1 |B.1.1.7	32.13949738	18.34766832	29.87837656	19.63445773
MT775558.1 |B.1.1.25	32.14321633	18.374736	29.86690804	19.61513963
MZ749908.1 |B.1.1	32.13699823	18.37021661	29.87378218	19.61900298
OM049448.1 |B.1.617.2	32.17268793	18.34557714	29.86146053	19.6202744
ON026035.1 |BA.1	32.1366201	18.35034927	29.88984417	19.62318646
ON026034.1 |BA.2	32.22110215	18.30981183	29.8655914	19.60349462

Total nucleotide composition: The total nucleotide composition was generated from the alignment and the nucleotide compositions were recorded.

The results show that the omicron variants have relatively greater percentage of guanine (G) and cytosine (C) contents in its sequence compared to the others. Since G and C forms a triple bond between them, it is relatively stronger than the adenine (A) and thymine (T) double bond. As a result, the G-C bond is harder to break that gives the virus more robustness than others. At the same time, there are less chances of mutations from them.

Phylogenetic tree: From the phylogenetic tree obtained by the Maximum Likelihood method shows that both the omicron variants that are almost entirely responsible for the recent infections in Bangladesh are not originated from the other variants like alpha or delta.

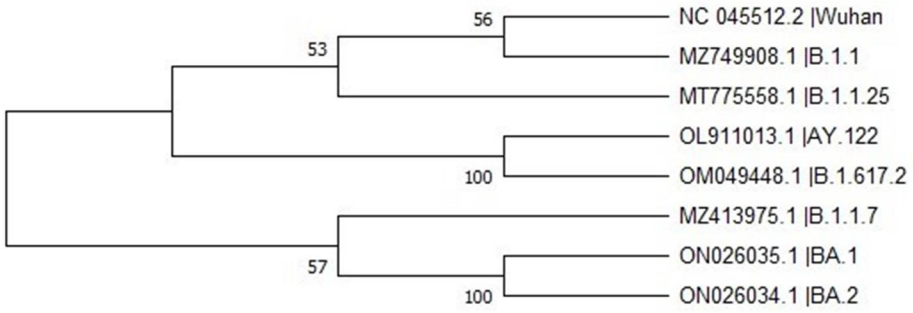

Figure 3. Phylogenetic tree of the dominant Covid-19 variants in Bangladesh using Maximum Likelihood method and Kimura-2 model.

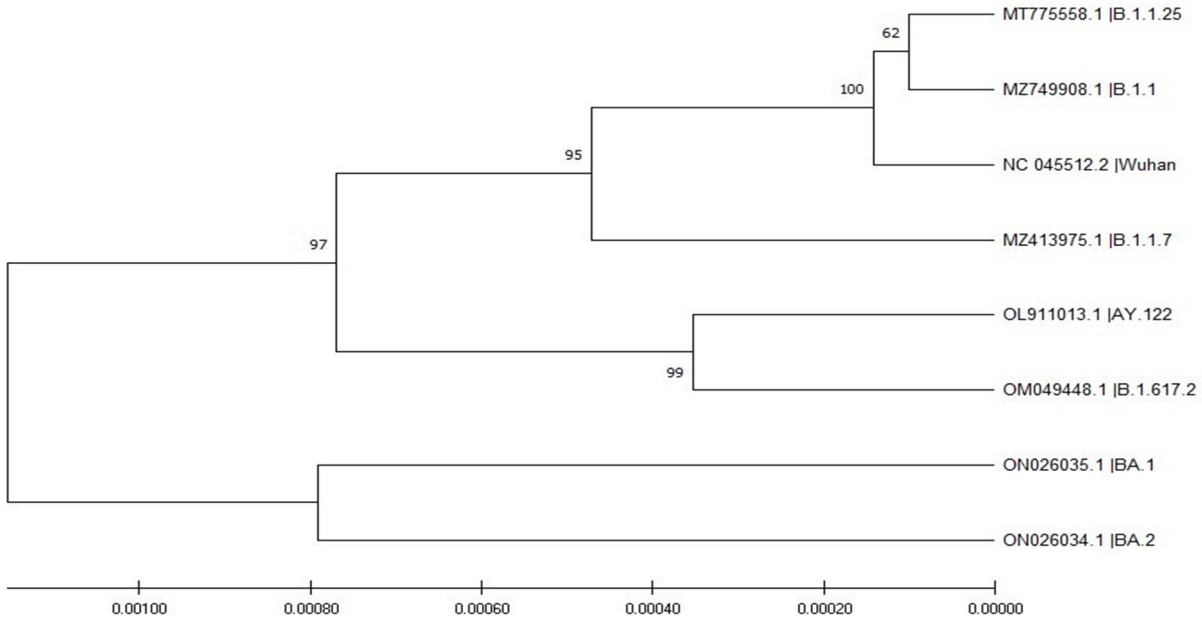

Figure 4. Phylogenetic tree of the dominant Covid-19 variants in Bangladesh using UPGMA method and Jukes-Cantor model.

The phylogeny analysis suggests that there is more possibility of the omicron to mutate again and another covid wave may be seen in near future. As phylogeny is a very important tool in microbial evolution and the treatment against them with new drug discovery and other conventional drugs, existing omicron lineages and others to come in near future cannot be treated as like the other variants of concern.

Protein mutation in spike region: The mutations in the spike protein are shown in table 5.

Table 5. Mutations in spike proteins with identity and coverage.

Variant	% Identity	% Coverage	Mutation in spike protein
B.1.1.25	99.8	100	L5F, D614G
B.1.1	99.9	100	D614G
B.1.617.2	99.5	99.8	T19R, F157del, R158del, L452R, T478K, D614G, P681R
B.1.1.7	99.4	99.8	H69del, V70del, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H, V1164A
AY.122	99.4	99.8	T19R, E156G, F157del, R158del, L452R, T478K, D614G, P681R, A879S, D950N
BA.2	97.9	99.8	T19I, L24del, P25del, P26del, A27S, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, A688V, N764K, D796Y, Q954H, N969K
BA.1	97.6	99.5	A67V, H69del, V70del, T95I, G142D, V143del, Y144del, Y145del, N211del, L212I, ins214EPE, G339D, S371L, L373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F

Analysis of infection data from European countries, where the S(G614) genotype is dominant, estimated the doubling time at approximately 3 days, significantly shorter than the 6-day doubling time of the initial outbreak in China. Another group observed a similar trend, although not statistically significant in their dataset (Jackson *et. al.*, 2021). It is suggested that viruses bearing S(G614) are 31% more transmissible than those bearing S(D614), (Guruprasad, 2021) although this analysis did not consider the effect of concomitant changes at other loci in the circulating viral strains. With D614G mutation in every one of them omicron (BA.1, BA.2) has the greatest number of changes followed by the B.1.1.7, AY.122, B.1.617.2. This makes the omicron as the most mutated variant of concern. The N501Y mutation increases the binding affinity with the ACE2 receptor, which is a major influencer of increased transmission, and in combination with Q498R, the binding affinity gets stronger and the Omicron variant gets easy access into the host (*CDC. Science Brief: Omicron (B.1.1.529) Variant. CDC. Accessed December 16, 2021.*, n.d.). P681H can multiply transmissibility by increasing the spike protein cleavage (Wu *et. al.*, 2021). Moreover, Q493R, N501Y, S371L, S373P, S375F, Q498R, and T478K are responsible for this higher affinity to the ACE2 (Jackson *et. al*., 2021).

Conclusion

The SARS-Cov-2 is continuously evolving, resulting in the introduction of several variants. Waves of severe infection are being caused by the variants of concerns, and they may continue with the new Omicron strain. Investigations are still being conducted to learn every detail about the SARS-Cov-2 Omicron variant in order to suggest effective measures to stop the impending surge.

Numerous COVID-19 variations that are currently circulating in various parts of Bangladesh show that Bangladeshi residents are extremely susceptible to this disease. We need a deeper knowledge of the pathobiology and evolution of this lethal virus in order to fight this disease.

Bangladesh is yet to vaccinate all of its citizen with the two doses of the vaccine. A relatively small number of people got the booster dose. While the omicron variant's infectivity and severity is still very much unclear, it is recommended to handle it with utmost care to prevent the possibility of an upcoming surge.

References

Alimohamadi, Y., Sepandi, M., Taghdir, M. and Hosamirudsari, H. 2020. Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis. *J. Prev. Med. Hyg.* **6**, E304-E312. https://doi.org/ 10.15167/2421-4248/jpmh2020.61.3.1530

CDC. Science Brief: Omicron (B.1.1.529) Variant. CDC. Accessed December 16, 2021. (n.d.).

Erles, K., Toomey, C., Brooks, H.W. and Brownlie, J. 2003. Detection of a group 2 coronavirus in dogs with canine infectious respiratory disease. *Virology* **310**, 216–223. https://doi.org/10.1016/S0042-6822(03)00160-0

Guruprasad, L. 2021. Human SARS-Cov-2 spike protein mutations. *Proteins: Structure, Function and Bioinformatics* **89**, 569–576. https://doi.org/ 10.1002/prot.26042

https://cov-lineages.org/lineage_list.html. (n.d.). Retrieved July 29, 2022, from https://cov-lineages.org/ lineage_list.html

Jackson, C.B., Zhang, L., Farzan, M. and Choe, H. 2021. Functional importance of the D614G mutation in the SARS-CoV-2 spike protein. *Biochem. Biophys. Res. Commun.* **538**, 108-115. https://doi.org/ 10.1016/j.bbrc.2020.11.026

Liya, G., Yuguang, W., Jian, L., Huaiping, Y., Xue, H., Jianwei, H., Jiaju, M., Youran, L., Chen, M. and Yiqing, J. 2020. Studies on viral pneumonia related to novel coronavirus SARS-Cov-2, SARS-Cov and MERS-Cov: a literature review. *APMIS* **128**, 423–432. https://doi.org/10.1111/apm.13047

Wu, L., Zhou, L., Mo, M., Li, Y., Han, J., Li, J.,Yang, Y., Zhang, X., Gong, C., Lu, K., Gong, L., Wu, C., Zhu, W. and Xu, Z. 2021. The effect of the multiple mutations in omicron RBD on its binding to human ACE2 receptor and immune evasion: an investigation of molecular dynamics simulations. ChemRxiv, Version 1, 2021.

Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., L.u, G., Wu, Y., Yan, J., Shi, Y., Zhang, X. and Gao, G. F. 2017. Cryo-EM structures of MERS-Cov and SARS-Cov spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **8**, 15092 https://doi.org/10.1038/ncomms15092

Zhou, P., Yang, X. Lou, Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. Di, Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., … Shi, Z. L. 2020. Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **588**, E6. https://doi.org/10.1038/s41586-020-2951-z