

Novel mutations in the spike protein of SARS CoV 2 from Bangladesh

Sultana Afrin¹, Jahan Begum², Syeda Nasreen¹, Salma Ahmed¹, Fahim Ahmad³, Md. Abdul Aziz⁴, Shyamal Paul⁵, and Rokshana Parvin²

¹Mymensingh Medical College

²Bangladesh Agricultural University

³TMSS Medical College

⁴Rangpur Medical College

⁵Netrokona Medical College

February 17, 2021

Abstract

Genetic characterization may provide useful insights into severe acute respiratory coronavirus 2 (SARS CoV 2) circulating in Bangladesh. Here we analyzed the SARS Cov 2 positive 41 nasopharyngeal samples collected in lysis buffer obtained at different regional national laboratories of Bangladesh during July – December 2020. Full length spike gene was amplified and SARS CoV 2 was confirmed upon Sanger sequencing. Multiple ClustalW alignment and phylogenetic study of 15 strains showed that genetically diverse SARS CoV 2 is circulating in the country where 80 documented and 55 novel substitutions have been observed in different regions of spike glycoprotein. Major mutations and/or deletions was identified at conserved amino acid positions that are functionally linked to host transition, antigenic drift, host surface receptor binding or antibody recognition sites, and viral oligomerization interfaces that may have significant effects on pathogenic capacity and epidemiological signatures.

INTRODUCTION

Corona Virus Disease 2019 (COVID-19) pandemic is caused by a novel species of the severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) (Peeri et al., 2020). The virus surfaced in Wuhan, China, in December 2019 and since then its worldwide spread impose a detrimental effect on global health and economy (Wang et al., 2020). In Bangladesh, the first case of COVID-19 has been identified on 8th March, 2020 (<https://corona.gov.bd/>). To prevent further human-to-human transmission, measures were scheduled quickly by the Government of Bangladesh including initial lockdown, social distancing, movement restrictions, closure of educational institutes, public and private offices, food outlets, factories, markets, and reduction of domestic and international air transport. Practical obstacles of putting these measures into strict actions in one of the world's most densely populated countries leaves Bangladesh vulnerable to the current COVID-19 pandemic. As a consequence, until 31st January 2021, a total of 535,230 cases with 8102 death have been recorded (<https://corona.gov.bd/?gclid>) officially in the country.

Given the evolving nature of the SARS-CoV-2 genome, drug and vaccine developers continue to be vigilant for the emergence of new variants or sub-strains of the virus (Nakagawa & Miyazawa, 2020). Knowledge of the molecular background of circulating strains is vital to select appropriately protective vaccine targets and to keep molecular diagnostic tools fit for purpose in Bangladesh. In the present study, amino acid substitutions in the spike (S) protein of selected SARS CoV 2 strains from Bangladesh were determined with the aim to recognize novel mutations that may further affect host-virus interactions and/or epidemiology.

Nasopharyngeal swabs in lysis buffer are being collected from COVID suspected patients in regional national laboratories and tested for SARS-CoV 2 by RT-qPCR targeting ORF1ab and N gene of the virus. Only SARS CoV 2 positive samples are being stored in the respective laboratory repositories. We received 41 samples from the repository of regional national laboratories which were collected during July – December 2020. Based on the viral load (C_q value of ORF 1ab and N gene), 15 samples were considered for further full-length amplification of Spike (S) gene by conventional RT-PCR (**Supplemental Tables S1 and S2**) and products were Sanger-sequenced. The obtained sequences are available at the GISAID database (accession numbers are listed in **Supplemental Table S3**).

The 15 S gene sequences obtained from Bangladeshi SARS CoV 2 patients fell into clade G (GISAID nomenclature) and further clustered with the B.1 lineage according to the dynamic lineage nomenclature (Rambaut et al., 2020). Bangladeshi S gene sequences showed the highest nucleotide homology (99.21-100%) to strains from USA, India, Bangladesh, Netherlands, Colombia, Egypt, Australia, South Africa and Canada (**Supplemental Table S4**,) indicating diverse group of virus circulation in the country. In the phylogenetic tree studied S gene is dispersed into many subbranches and remained in closer with their respective similar strains from different countries (**Figure 1**). Newly characterized S genes showed heterogeneity among themselves and were kept distant from one another in the tree indicated probable evolution of the circulating viruses.

The S glycoprotein consists of two functional subunits, S1 and S2 separated by furin cleavage site (FCS) is crucial for viral infection, has been undergoing mutations and is highly glycosylated (Li et al., 2020; Conceicao et al., 2020; Shang et al., 2020). The analysis of deduced amino acid (aa) sequences of the 15 strains exhibited a total of 135 replacements (substitution, deletion or insertion) of which 80 replacements have been described previously and 55 were novel in these Bangladeshi strains (**Table 1**). In order to characterize these new replacements, we used the CoVsurver algorithm available at the GISAID platform. Replacement was distributed all over the S glycoprotein however; only functional mutations are critically important to investigate the biological significance of these mutations (Supplemental Table S5). The novel Y144del, E298stop replacement at N-terminal domain along with substitutions S359I, G381A, P491H, G496W, P499A and L518V in the RBD regions in studied Bangladeshi strains (Figure 2a) may affect in receptor binding and/or form parts of neutralization-relevant epitopes (Andersen et al., 2020). Although the strains with amino acid changes at RBD studied here may not likely become widespread, but suggesting this RBD natural variants with increased resistance to antibody-mediated neutralization should be closely watched. SARS-CoV-2 variants with S protein D614G mutations now predominate globally. A number of characteristic variants and combined variants with D614G substitution across S glycoprotein have been identified previously and has described as modifying the pattern of infection (Li et al., 2020). Of which, D614G+Q675H variant has been observed in M77 and K6 Bangladeshi strains and are distributed in many other countries. In addition, another combined variant D614G+L5F replaced in Bangladeshi strain as a novel combination of D614G+L5V (Figure 2a), must be further investigated for its infectivity. Furin cleavage site (FCS) is cleaved by furin proteases mediates the entrance of SARS-CoV-2 into host cells and expand the virus tropism and also has significantly contributed in new variants of SARS CoV 2 reported in UK and Brazil (Walls et al., 2020). Herein, we detected well established P681R substitution at FCS in D3 strain. Furthermore, two novel substitutions N282D and N1098D found in D8 and D3 respectively, removes a potential N-glycosylation site which may affect antigenic property and have implications in viral pathobiology and vaccine design (Xu et al., 2020; Watanabe et al., 2020). Altogether, three novel mutations at RBD (P491H, P499A, S514C), along with presence of well-established variant D614G, P681R substitution at FCS and novel removal of N-glycosylation site N1098D made D3 a significant strain among 15 characterized Bangladeshi strains is depicted on crystal structure (Figure 2b).

While evidence of different novel mutations is still accumulating, the currently notifying novel spike protein mutations in SARS CoV 2 in Bangladesh gives rise to the possibility of a fitness advantage, which is likely to allow more successful person-to-person transmission. The impact of such novel mutation needs to be investigated further.

In conclusion, the research detected known and novel mutations in the spike glycoprotein in 15 SARS CoV 2 strains from Bangladesh. Due to limited resources, the study of the whole genome was not feasible, but could provide valuable insights into other structural and accessory proteins. However, in the event of COVID-19 pandemic this study will add to a basis required for understanding the circulation of SARS-CoV-2 in Bangladesh.

ACKNOWLEDGEMENTS

The authors are grateful to the district laboratories for sending sample to the studied laboratory. We acknowledge Mr. Shafiqul Islam for his excellent technical support throughout the work. We also acknowledge Sultana Shabnam Nila and Jobyda Khanam for their assistant at sample collection. This research was supported by the project funded by The World Academy of Science (TWAS), grant number No. 20-284 RG/BIO/AS_G – FR3240314166 and return fellowship by Alexander Von Humboldt Foundation, Germany.

ETHICAL STATEMENTS

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. This work involved handling of samples that are already being deposited on the respective laboratory repositories. Written permission from the Institutional Review Board (IRB) was taken to use the sample for further analysis (MMC/IRB/2020/290).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY

The data and the sequence accession numbers are listed in the supplementary materials of this article

REFERENCES

- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature medicine*, 26(4), 450–452. <https://doi.org/10.1038/s41591-020-0820-9>
- Conceicao, C., Thakur, N., Human, S., Kelly, J. T., Logan, L., Bialy, D., Bhat, S., Stevenson-Leggett, P., Zagrajek, A. K., Hollinghurst, P., Varga, M., Tsirigoti, C., Tully, M., Chiu, C., Moffat, K., Silesian, A. P., Hammond, J. A., Maier, H. J., Bickerton, E., Shelton, H., ... Bailey, D. (2020). The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS biology*, 18(12), e3001016. <https://doi.org/10.1371/journal.pbio.3001016>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., Qin, H., Wang, M., Lu, Q., Li, X., Sun, Q., Liu, J., Zhang, L., Li, X., Huang, W., & Wang, Y. (2020). The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell*, 182(5), 1284–1294.e9. <https://doi.org/10.1016/j.cell.2020.07.012>
- Nakagawa, S., & Miyazawa, T. (2020). Genome evolution of SARS-CoV-2 and its virological characteristics. *Inflammation and regeneration*, 40, 17. <https://doi.org/10.1186/s41232-020-00126-7>
- Peeri, N. C., Shrestha, N., Rahman, M. S., Zaki, R., Tan, Z., Bibi, S., Baghbanzadeh, M., Aghamohammadi, N., Zhang, W., & Haque, U. (2020). The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned?. *International journal of epidemiology*, 49(3), 717–726. <https://doi.org/10.1093/ije/dyaa033>
- Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature*

microbiology, 5(11), 1403–1407. <https://doi.org/10.1038/s41564-020-0770-5>

Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., & Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America*, 117(21), 11727–11734. <https://doi.org/10.1073/pnas.2003138117#>

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, 10(3), 512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Velesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181(2), 281–292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>

Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *Lancet (London, England)*, 395(10223), 470–473. [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9)

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., & Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science (New York, N.Y.)*, 369(6501), 330–333. <https://doi.org/10.1126/science.abb9983>

Xu, W., Wang, M., Yu, D., & Zhang, X. (2020). Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles During Global Transmission Course of COVID-19. *Frontiers in immunology*, 11, 565278. <https://doi.org/10.3389/fimmu.2020.565278>

Figure Legends

Figure 1. Phylogenetic tree based on Spike gene (3778 bp) of SARS CoV 2 of 15 Bangladeshi strains. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [12]. This analysis involved 65 nucleotide sequences with 1000 bootstrap replication. Evolutionary analyses were conducted in MEGA X [13]. Strains from the present study are highlighted in navy color closed square box and taxon name. Other Bangladeshi strains are shown in closed circle (pink color). The tree is rooted to the Wuhan-hu-1 strain. Bootstrap values are shown in the branch node and variation scale is described at the bottom.

Figure 2. Structure of spike glycoprotein of Bangladeshi SARS CoV 2 strain. a) Genomic sketch of spike protein (1-1273aa) with highlighted close view of FCS region (680-689) of all studied strains. Novel replacements and known variants are shown in the genomic sketch b) Representation of chain and cartoon view of crystal structure of D3 strain along with depicted novel and established substitutions. The crystal structure was created using the online based modeling tool called Swiss-Model (<https://swissmodel.expasy.org>).

Supplementary materials

Table S1: Distribution and viral load of selected SARS CoV 2 positive sample used in the study

Table S2: The newly designed primer for amplification of S gene of SARS CoV-2 in reference to EPI_ISL_487367

Table S3: Strains name and accession numbers of the Spike gene submitted to GISAID

Table S4: Relatedness of studied strains with other most closely related isolates/strains

Table S5: List of functional mutations of the Spike protein noticed in Bangladeshi SARS CoV-2 strains and its global occurrence

Hosted file

Table 1.pdf available at <https://authorea.com/users/396267/articles/509440-novel-mutations-in-the-spike-protein-of-sars-cov-2-from-bangladesh>



