

Modest Evolutionary Changes of the SARS-CoV-2 Genome in Bangladesh

Sezanur Rahman

International Centre for Diarrhoeal Disease Research, Bangladesh <https://orcid.org/0000-0002-0092-9261>

Mehedi Hasan

ICDDR B: International Centre for Diarrhoeal Disease Research Bangladesh

Mohammad Enayet Hossain

ICDDR B: International Centre for Diarrhoeal Disease Research Bangladesh

Mohammed Ziaur Rahman

ICDDR B: International Centre for Diarrhoeal Disease Research Bangladesh

Mustafizur Rahman (✉ mustafizur@icddr.org)

Virology Laboratory, Infectious Diseases Division, icddr,b, Mohakhali, Dhaka, <https://orcid.org/0000-0002-6876-0191>

Research article

Keywords: SARS-CoV-2, mutation rate, evolution, Bangladesh, South Asia

Posted Date: December 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-122444/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The human-to-human transmissible nature of SARS-CoV-2 makes Bangladesh, as well as the other South Asian regions, vulnerable to the ongoing pandemic of COVID-19 due to their high population densities. The present study was designed based on the genome wide analysis of Bangladeshi and other South Asian isolates. Complete sequences of SARS-CoV-2 were retrieved from the EpiCoV database in order to identify molecular features demonstrating the evolutionary trail and mutation rate.

Result: The complete genome mutation rate of the Bangladeshi isolates was estimated to be $0.49E-3$ nucleotide substitutions/site/year. A higher mutation rate was found in the non-structural protein-coding genes at: ORF6 ($10.29E-3$), ORF7a ($31.81E-3$), and ORF8 ($18.35E-3$). In contrast, the mutation rates of the structural protein-coding genes were relatively low at: M ($1.14E-3$), S ($1.47E-3$), E ($3.35E-3$), and N ($4.59E-3$).

Conclusions: A comparison of Bangladeshi and other South Asian isolates demonstrated that there were limited mutational changes in the SARS CoV-2 genome. Knowledge of the Southeast Asian SARS CoV-2 evolutionary genome will help in selecting future vaccine candidates and designing therapeutic drug targets.

Background

A novel β -coronavirus nCoV-19, later known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), caused acute pneumonia in Wuhan City, China on December 2019 (1). SARS-CoV-2 has rapidly evolved into a global pandemic. Just in one year of its emergence, the virus has infected 56.3 million people and is responsible for more than 1.3 million deaths worldwide (2).

The human-to-human transmissible nature of SARS-CoV-2 makes South Asia, the most densely populated region in the world, extremely vulnerable to the current pandemic. The first laboratory-confirmed case in Southeast Asia was reported in Nepal, which was documented on January 23, 2020 (3). This initial report was followed by case reports from Sri Lanka, India, Pakistan, Bhutan, Maldives and Bangladesh. In terms of the number of coronavirus cases and deaths, India was the first South Asian country to overtake China, followed by Pakistan and Bangladesh. India is the world's second-worst affected country with total cases of more than 6 million, trailing only the United States, which has more than 7 million cases (2). In Pakistan, cases have fallen from a peak in mid-June of almost 6,000 new infections, to about a few hundred a day in September. Bangladesh, which had a total of 438,795 cases as of 18 November, saw its daily cases peak in mid-June before they dropped in the last week of July. Nepal has had a relatively low overall number of cases but is now seeing their reported infection rate rising again after an initial peak in June. Sri Lanka has much lower levels of infection t (2).

SARS-CoV-2, a positive-sense single-stranded RNA virus, has a genome size of approximately 30 kb which encodes 27 proteins from 11 open reading frames (ORF) (4). The largest ORF (ORF1ab) constitutes about 70% of the genome and encodes polyproteins PP1ab and PP1a (5). These two polyproteins are

further cleaved into 16 non-structural proteins (NSP1 to 16). The NSPs play a role in pathogenesis and are responsible for replication, transcription, proteolytic processing (6-8). The RNA-dependent polymerase is a potential antiviral drug target (9, 10). The remaining ten ORFs, encode four structural proteins and at least six accessory proteins. The structural proteins are spike glycoprotein (S), matrix protein (M), envelope protein (E), and nucleocapsid protein (N), while the accessory proteins are ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 (4, 11). The S protein is crucial for viral attachment and entry through ACE2 receptor recognition (12-14). Antibodies against the S protein were found to be effective in inhibiting SARS-CoV entry into the host cell (15) thus serving as a potential drug target and the basis of a future vaccine (16). On the other hand, mutations in S protein help the virus escape the host cell immune system and lead to the emergence of variant strains. As the most abundant structural protein, M protein can also be considered as an antiviral drug target (17). Data on the mutation rate can provide information about the level of genetic diversity in the population. Therefore, understanding the evolution of SARS-CoV-2 at the genomic and individual gene level is important for the development of low cost diagnostic tools, therapeutics and vaccines.

In the present study, we analysed the full genome of SARS-CoV-2 Bangladeshi isolates in order to geographical variants and the mutation rate. The evolutionary pattern of three major genes (ORF1a, S, and M), which are considered as drug/vaccine targets, were investigated by comparing Bangladeshi isolates with South Asian isolates and the reference Wuhan-Hu-1 strain.

Results

Geographic distribution of Bangladeshi isolates

A total of 306 SARS-CoV-2 full genomes were submitted from patients representing 20 different districts of Bangladesh (Figure1). The highest number (n=66) of was from the capital city Dhaka, while 20 samples were not specified at the district level.

The evolutionary rate for Bangladeshi isolates

We analysed four structural genes (S, E, M and N) and four major open reading frames (ORF1a, ORF6, ORF7a and ORF8) along with the full genome (Table-1). The mutation rate of the in full Bangladesh genome was $0.49E-3$ (95% HPD $0.30E-3$ to $0.68E-3$) nucleotide substitutions/site/year. In ORFs, the highest mutation rate was found for ORF7a ($31.8E-3$ nucleotide substitutions/site/year) and lowest for ORF1a ($0.59E-3$ nucleotide substitutions/site/year) while all segments showed negative or stabilizing selection (dN/dS, 0.47 to 0.95). In the structural proteins, the highest mutation rate was $4.59E-3$ nucleotide substitutions/site/year for N and it gradually decreased for E, S and M ($3.35E-3$, $1.47E-3$ and $1.14E-3$ nucleotide substitutions/site/year). Only the envelop protein gene (E) showed positive selection (dN/dS, 1.43) while other structural proteins showed negative selection (dN/dS, 0.20 to 0.57).

Mutation in S gene of Bangladeshi isolates

A total of 107 nucleotide changes were found in the Bangladeshi S gene; among them 53 were non-synonymous. Among 107 substitutions, 98 were present in <1% isolates, and six changes were present in <2% isolates (Supplementary Table 4). Nucleotide changes C13T, C882T and A1841G were present among 2.6%, 6.3% and 97.7% of isolates respectively. Among them C882T was synonymous, while C13T and A1841G substituted amino acid L5F and D614G. A non-synonymous change in amino acid position five (L5F) occurred between the same group of amino acids (non-polar Leucine to Phenylalanine). While, in position 614, a hydrophilic amino acid D (Aspartate) was substituted by an amphoteric amino acid, G (Glycine).

The evolutionary rate of S gene in South-Asia

The evolutionary rate of S gene varied in the different regions between $1.34E-3$ to $1.84E-3$ nucleotide substitutions/site/year, except for Maharashtra which was $3.50E-3$ (Table-2). With regards to the degree of natural selection (dN/dS) acting on all S genes, it ranged from 0.41 to 0.76 (stabilizing selection) except for Gujrat at 1.0 (neutral selection). The major substitution found for the Bangladeshi S gene (e.g., D614G substitution) was also analysed for Indian isolates. Only 15.7% of isolates from New Delhi contained this substitution and the highest substitution rate was identified for Gujarat isolates (93.8%) which are comparable with Bangladeshi isolates (97.7%).

Mean p-distance of genes from the reference strain Wuhan-Hu-1

Finally, from 'Dataset-2', three major genes (ORF1a, S, and M) were subjected to genetic distance analysis over time through the p-distance method. For Bangladeshi isolates, mean p-distance from the reference sequence (Wuhan-Hu-1) increased during first month then ran horizontally for ORF1a while it slightly increased for S gene (Figure-2). For M gene, p-distance showed a curved line starting with high similarity with Wuhan-Hu-1 strain, then increased and finally declined again. A similar pattern to Bangladesh was observed for Maharashtra and West Bengal for S and M genes. All Indian regions showed a slightly upward trend for ORF1a. Only Delhi and Gujarat showed an upward trend for S and M genes. Notably, in the last two time-points of these two regions, the numbers of isolates were less than ten (Supplementary Table-3).

Discussion

The study presented the evolutionary characteristics of SARS-CoV-2 circulating in Bangladesh and the South Asian region (mainly four states of India by sequence analysis). Here we provided: (i) the mutation rate of the SARS-CoV-2 genome in Bangladesh including the major genes (e.g., ORF1a, ORF6, ORF7a, ORF8, S, E, M and N); (ii) the type of substitution occurring in the S gene of Bangladeshi and South Asian isolates; (iii) the mutation rate of S gene from South Asian region; (iv) and the genetic changes over time in three longer genes (ORF1a, S, and M) from South Asian region through the p-distance methodology.

During the first global epidemic of SARS-CoV in 2003, the mutation rate was estimated to be $0.80E-3$ to $2.38E-3$ nucleotide substitutions/site/year (16). For MERS-CoV which caused an epidemic in the Middle

East in 2012, the mutation rate was 0.46E-3 to 1.12E-3 nucleotide substitutions/site/year (18, 19). In the early stage of current pandemic, several studies estimated the mutation rate of full genome of SARS-CoV-2 to be approximately 1.05E-3 to 1.80E-3 nucleotide substitutions/site/year (12, 20) which is comparable with our result (0.49E-3 nucleotide substitutions/site/year) for Bangladesh. For Bangladeshi isolates, ORF7a, ORF8 and ORF6 showed higher mutation rates compared to other genes (N, E, S, M and ORF1a). However, all mutations were conserved during evolution (negative selection, dN/dS value <1) except the envelop protein gene (E) which adapted (dN/dS=1.43). Importantly, the mutation was conserved during the evolution of the spike protein gene (S).

Amino acid position 614 in the S gene is located at the C-terminal region of the S1 domain and does not belong to a receptor binding domain. This study showed that 98% of the Bangladeshi isolates had D614G substitution. This mutation received attention because some studies claimed that the D614G mutation might be involved in increased infectivity and mortality (21-23). Moreover, one study claimed that the substitution was most likely neutral due to protein function and interaction with the human ACE2 receptor by *in silico* analysis (24) and another clinical study supported this finding where no association of this mutation was found with severe illness (25). Although a high D614G substitution was present in the Bangladeshi isolates, the case fatality rate (1.41%) was low compared to other regions [i.e. West Bengal (1.94%), Delhi (2.10%), Maharashtra (2.75%) and Gujarat (2.77%)] (2) (Table-2).

The overall mutation rate for the S gene of SARS-CoV-2 is lower than other respiratory RNA virus membrane glycoprotein genes or other genes used for drug and vaccine targets. For example, the HA and NA gene of influenza virus have mutation rates of 2.0E-3 to 10.0E-3 nucleotide substitutions/site/year (26-29), the SH and G gene of Respiratory Syncytial Virus have rates of 2.5E-3 to 3.49E-3 nucleotide substitutions/site/year (30, 31), and human metapneumovirus rates of 1.6E-3 to 3.61E-3 nucleotide substitutions/site/year (31, 32).

This study aimed to cover all of the South Asian region, but due to a low number of the available sequences, we had to exclude Pakistan, Sri-Lanka, Nepal, the Maldives, Bhutan and some states of India (Supplementary Table-2). Here we examined how average genetic distances of SARS-CoV-2 isolates from different regions changed over time. In our analysis, we did not show an upward trend of genetic changes except some fluctuations due to the low numbers of isolates during that time points. This indicates that the virus is not evolving away from Wuhan-Hu-1 strain significantly as reported previously (33). The only exception was for the M gene of the Gujarat isolates with upward trends, indicating a gene-specific evolutionary change.

Conclusions

A comparison of the Bangladeshi and other South Asian isolates with the Wuhan-Hu-1 strain provides evidence that that SARS-CoV-1 genome is largely conserved and this has important implications for future vaccine and drug development.

Methods

Preparation of 'Dataset-1' for Bangladeshi isolates

In this study the full genomes of SARS-CoV-2 submitted from Bangladesh were retrieved from the EpiCoV database available via the GISAID website (www.gisaid.org) for analysis of evolutionary pattern. Multiple sequence alignment was performed using the MUSCLE programme (34) and gene segments (ORF1a, ORF6, ORF7a, ORF8, S, E, M and N) were segregated from align file with BioEdit (version 7.0.9.0). Geographical locations in district level were also retrieved (Supplementary Table-1) to prepare a sample distribution map.

Evolutionary rate and selection pressure measurement

An uncorrelated relaxed clock log-normal molecular clock model was used to calculate the evolutionary rate (numbers of nucleotide substitutions/site/year) with coalescent Bayesian skyline prior of the study isolates through a Markov Chain Monte Carlo (MCMC) framework in BEAST (v2.6.3)(35, 36). MCMC chains were run for 100 million steps with sampling every 10,000 steps from the posterior distribution to ensure adequate mixing of model parameters. Hasegawa-Kishino-Yano (HKY) nucleotide substitution model was used (12), and the evolutionary rate was evaluated using Tracer v1.7.1. Finally, the S gene of the study isolates were subjected for selection pressures calculate using Datamonkey server (37), by calculating the ratio of non-synonymous (dN), and synonymous (dS) nucleotide substitutions per site (dN/dS) (38).

Preparation of 'Dataset-2' for genetic distance analysis

To observe genetic change over time different datasets of sequences were prepared where the Wuhan-Hu-1 strain (NC_045512.2) was used as a reference sequence. For this dataset the full genomes of SARS-CoV-2 from South-Asian countries with at-least 100 isolates were retrieved from the EpiCoV database up to 32 weeks (January 01 to August 11, 2020) (Supplementary Table-2). During sequence retrieval, only Bangladesh and seven states of India (Delhi, Gujrat, Karnataka, Odisha, Telangana and West Bengal) having at-least 100 isolates were sampled during weeks of 11-32 (11 March to 11 August) of 2020. Finally, datasets of sequences were prepared for the individual region with two weeks' time-points. Datasets having a minimum of ten isolates in continuing five time-points were selected for analysis (Supplementary Table-3). Sequences which did not mention collection dates or had high ambiguity (contained a lot of Ns for example) were excluded during the preparation of the dataset. Finally, a total of 1480 isolates were included in the analysis from five South Asian regions; Bangladesh (n=306), Delhi (n=253), Gujarat (n=457), Maharashtra (n=312) and West Bengal (n=152).

Genetic distance matrix calculations

Multiple sequence alignment was performed using the MUSCLE programme (34) and MEGA-X software was used to calculate the distance matrix of study sequences by p-distance model (39). Genetic

distances were calculated between all possible pairs. By using distance matrix data the average distance between reference sequences and study sequences were calculated.

Abbreviations

SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus-2
COVID-19	coronavirus disease 2019
ORF	open reading frames
NSP	non-structural proteins
ACE	Angiotensin-Converting Enzyme
HPD	Highest Posterior Density
MCMC	Markov Chain Monte Carlo
HKY	Hasegawa-Kishino-Yano

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests

Funding

The manuscript is a part of 'COVID-19 testing and tracing in Bangladesh' study, which was funded by 'Bill and Melinda Gates Foundation' (INV-017556). The funding body had no role in designing the study protocol, the collection, analysis, interpretation of data, and the writing of the manuscript.

Authors' contributions

MR and MZR conceived and coordinated the study. SR and MH performed data set preparation, data analysis, interpretation of the results and wrote the manuscript. MEH involved in critically reviewing the manuscript and provided intellectual input. All authors reviewed subsequent drafts of the manuscript and approved the final version.

Acknowledgements

The manuscript is a part of 'COVID-19 testing and tracing in Bangladesh' study, which was funded by 'Bill and Melinda Gates Foundation'. icddr,b acknowledges with gratitude the commitment of 'Bill and Melinda Gates Foundation' to its research efforts. icddr,b is also grateful to the Governments of Bangladesh, Canada, Sweden and the UK for providing core/unrestricted support.

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *The Lancet*. 2020;395(10223):470-3.
2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*. 2020;20(5):533-4.
3. Bastola A, Sah R, Rodriguez-Morales AJ, Lal BK, Jha R, Ojha HC, et al. The first 2019 novel coronavirus case in Nepal. *The Lancet Infectious Diseases*. 2020;20(3):279-80.
4. Helmy YA, Fawzy M, Elswad A, Sobieh A, Kenney SP, Shehata AA. The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *Journal of Clinical Medicine*. 2020;9(4):1225.
5. Baranov PV, Henderson CM, Anderson CB, Gesteland RF, Atkins JF, Howard MT. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology*. 2005;332(2):498-510.
6. Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: the role of the nsp2 and nsp3 in its pathogenesis. *Journal of medical virology*. 2020;92(6):584-8.
7. Peng Q, Peng R, Yuan B, Zhao J, Wang M, Wang X, et al. Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. *Cell reports*. 2020;31(11):107774.
8. Wang Q, Wu J, Wang H, Gao Y, Liu Q, Mu A, et al. Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell*. 2020;182(2):417-28. e13.
9. Yin W, Mao C, Luan X, Shen D-D, Shen Q, Su H, et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*. 2020.
10. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 2020:1-13.

11. Xu J, Hu J, Wang J, Han Y, Hu Y, Wen J, et al. Genome organization of the SARS-CoV. *Genomics, proteomics & bioinformatics*. 2003;1(3):226-35.
12. Li X, Wang W, Zhao X, Zai J, Zhao Q, Li Y, et al. Transmission dynamics and evolutionary history of 2019-nCoV. *J Med Virol*. 2020;92(5):501-11.
13. Ortega JT, Serrano ML, Pujol FH, Rangel HR. Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis. *EXCLI journal*. 2020;19:410-7.
14. Wu Y. Strong evolutionary convergence of receptor-binding protein spike between COVID-19 and SARS-related coronaviruses. *bioRxiv*. 2020.
15. Salvatori G, Luberto L, Maffei M, Aurisicchio L, Roscilli G, Palombo F, et al. SARS-CoV-2 SPIKE PROTEIN: an optimal immunological target for vaccines. *Journal of Translational Medicine*. 2020;18:1-3.
16. Zhao Z, Li H, Wu X, Zhong Y, Zhang K, Zhang Y-P, et al. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC evolutionary biology*. 2004;4(1):21.
17. Bianchi M, Benvenuto D, Giovanetti M, Angeletti S, Ciccozzi M, Pascarella S. Sars-CoV-2 Envelope and Membrane Proteins: Structural Differences Linked to Virus Characteristics? *BioMed Research International*. 2020;2020.
18. Cotten M, Watson SJ, Zumla AI, Makhdoom HQ, Palser AL, Ong SH, et al. Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *MBio*. 2014;5(1).
19. Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel-human interface. *Elife*. 2018;7:e31257.
20. Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. *Bulletin of the World Health Organization*. 2020;98(7):495.
21. Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *International Journal of Clinical Practice*. 2020.
22. Hu J, He CL, Gao Q, Zhang GJ, Cao XX, Long QX, et al. The D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity. *BioRxiv*. 2020.
23. 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.
24. Isabel S, Grana-Miraglia L, Gutierrez JM, Bundalovic-Torma C, Groves HE, Isabel MR, et al. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *bioRxiv*. 2020.
25. Mak GCK, Lau AWL, Chan AMY, Chan DYW, Tsang DNC. The D614G substitution in the S gene and clinical information for patients with COVID-19 detected in Hong Kong. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2020;130:104550.
26. Mugosa B, Vujosevic D, Ciccozzi M, Valli MB, Capobianchi MR, Lo Presti A, et al. Genetic diversity of the haemagglutinin (HA) of human influenza a (H1N1) virus in montenegro: Focus on its origin and evolution. *J Med Virol*. 2016;88(11):1905-13.

27. Qiu X, Bahl J. Structurally informed evolutionary models improve phylogenetic reconstruction for emerging, seasonal, and pandemic influenza viruses. *bioRxiv*. 2019:228692.
28. Youk S, Lee D-H, Ferreira HL, Afonso CL, Absalon AE, Swayne DE, et al. Rapid evolution of Mexican H7N3 highly pathogenic avian influenza viruses in poultry. *PLoS One*. 2019;14(9):e0222457.
29. Kwon J-H, Lee D-H, Criado MF, Killmaster L, Ali MZ, Giasuddin M, et al. Genetic Evolution and Transmission Dynamics of Clade 2.3. 2.1 a Highly Pathogenic Avian Influenza A/H5N1 Viruses in Bangladesh. *Virus Evolution*. 2020.
30. Haider MSH, Deeba F, Khan WH, Naqvi IH, Ali S, Ahmed A, et al. Global distribution of NA1 genotype of respiratory syncytial virus and its evolutionary dynamics assessed from the past 11 years. *Infection, Genetics and Evolution*. 2018;60:140-50.
31. Jagusic M, Slovic A, Ivancic-Jelecki J, Ljubin-Sternak S, Vilibić-Čavlek T, Tabain I, et al. Molecular epidemiology of human respiratory syncytial virus and human metapneumovirus in hospitalized children with acute respiratory infections in Croatia, 2014–2017. *Infection, Genetics and Evolution*. 2019;76:104039.
32. Kim JI, Park S, Lee I, Park KS, Kwak EJ, Moon KM, et al. Genome-wide analysis of human metapneumovirus evolution. *PLoS one*. 2016;11(4):e0152962.
33. Genomic diversity and divergence of SARS-CoV-2/COVID-19 [Internet]. *observablehq.com*. 2020. Available from: <https://observablehq.com/@spond/current-state-of-sars-cov-2-evolution>.
34. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*. 2004;32(5):1792-7.
35. Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. *Bioinformatics*. 2009;25(11):1370-6.
36. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology*. 2019;15(4):e1006650.
37. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Molecular biology and evolution*. 2018;35(3):773-7.
38. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution*. 2005;22(5):1208-22.
39. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*. 2018;35(6):1547-9.

Tables

Table 1: Evolutionary rate of Bangladeshi SARS-CoV-2 isolates

¹ Selection pressures calculate using Datamonkey server, by calculating the ratio of non-synonymous (dN), and synonymous (dS) nucleotide substitutions per site.

Gene name	Mutation rate in nucleotide substitutions/site/year (95% HPD)	dN/dS ¹
S	1.47×10^{-3} ($0.87 \times 10^{-3} - 2.05 \times 10^{-3}$)	0.41
E	3.35×10^{-3} ($0.07 \times 10^{-3} - 9.92 \times 10^{-3}$)	1.43
M	1.14×10^{-3} ($0.30 \times 10^{-3} - 2.29 \times 10^{-3}$)	0.20
N	4.59×10^{-3} ($2.28 \times 10^{-3} - 7.53 \times 10^{-3}$)	0.57
ORF1a	0.59×10^{-3} ($0.39 \times 10^{-3} - 0.79 \times 10^{-3}$)	0.49
ORF6	10.29×10^{-3} ($1.48 \times 10^{-3} - 24.80 \times 10^{-3}$)	0.95
ORF7a	31.81×10^{-3} ($4.87 \times 10^{-3} - 64.30 \times 10^{-3}$)	0.47
ORF8	18.35×10^{-3} ($6.27 \times 10^{-3} - 31.30 \times 10^{-3}$)	0.75
Full genome	0.49×10^{-3} ($0.30 \times 10^{-3} - 0.68 \times 10^{-3}$)	-

Table 2: Evolutionary rate and presences of D614G substitution of S gene of SARS-CoV-2 from Bangladesh and different provinces of India

¹ Selection pressures calculate using Datamonkey server, by calculating the ratio of non-synonymous (dN), and synonymous (dS) nucleotide substitutions per site.

²CFR (Case Fatality Rate) was retrieved from database of Johns Hopkins Coronavirus Resource Center on September 15, 2020 (Dong, Du et al. 2020).

Country/Region	No. of sequences analyzed	Mutation rate in nucleotide substitutions/site/year (95% HPD)	dN/dS ¹	Presence of D614G in percentage	CFR ²
Bangladesh	306	1.47×10^{-3} (0.87×10^{-3} – 2.05×10^{-3})	0.41	97.69	1.41
Delhi	253	1.34×10^{-3} (0.64×10^{-3} – 2.14×10^{-3})	0.76	15.65	2.10
Gujarat	457	1.53×10^{-3} (1.01×10^{-3} – 2.29×10^{-3})	1.00	93.87	2.77
Maharashtra	312	3.50×10^{-3} (2.20×10^{-3} – 4.91×10^{-3})	0.52	86.90	2.75
West Bengal	152	1.84×10^{-3} (1.02×10^{-3} – 2.78×10^{-3})	0.49	85.99	1.94

Figures

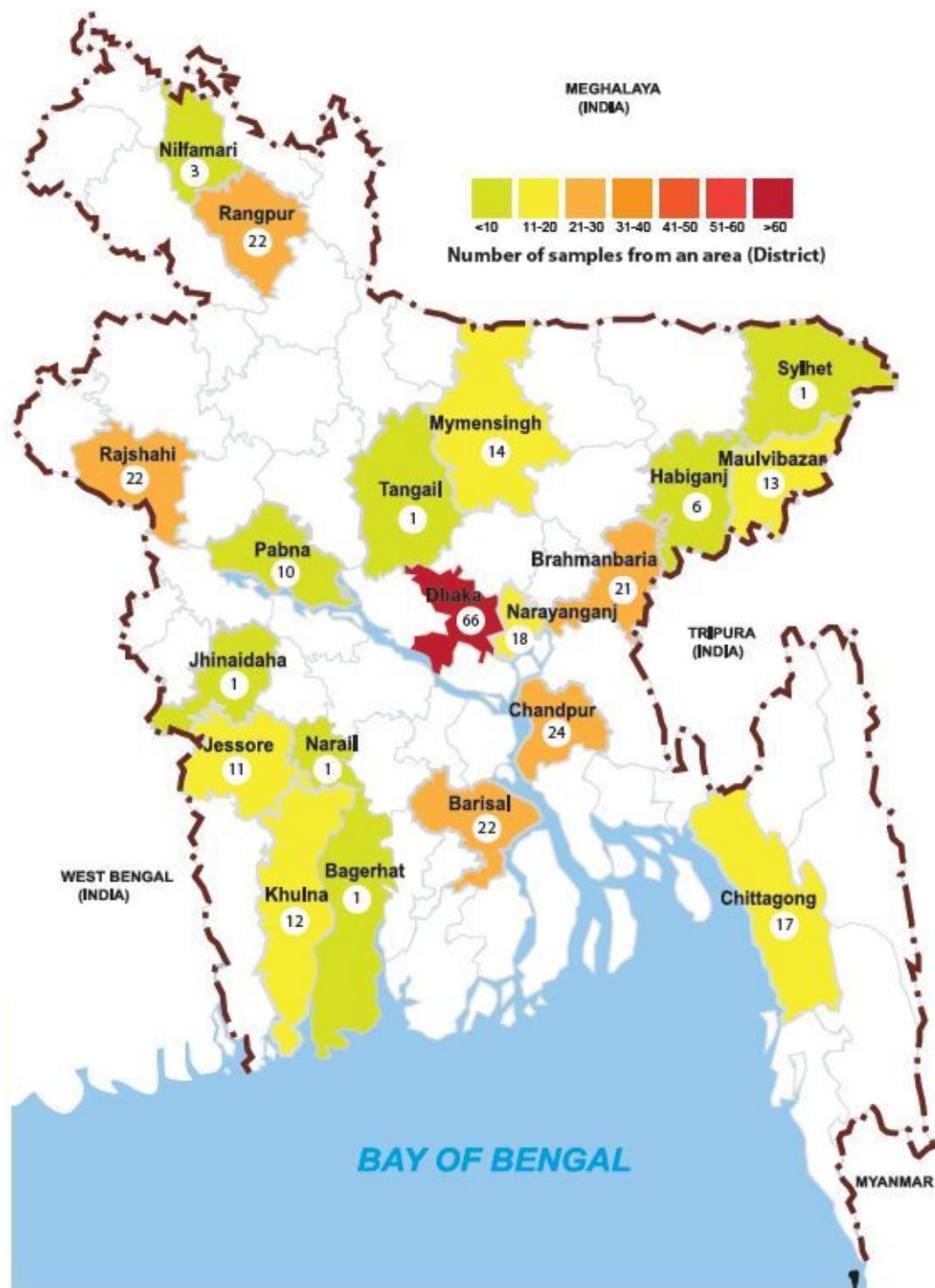


Figure 1

Available full genome sequence of SARS-CoV-2 from different geographical locations of Bangladesh (August 11, 2020). An open access geographical index map of Bangladesh was downloaded from the national portal of 'Geological survey of Bangladesh (www.gsb.gov.bd)' and re-created for the manuscript. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any

country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

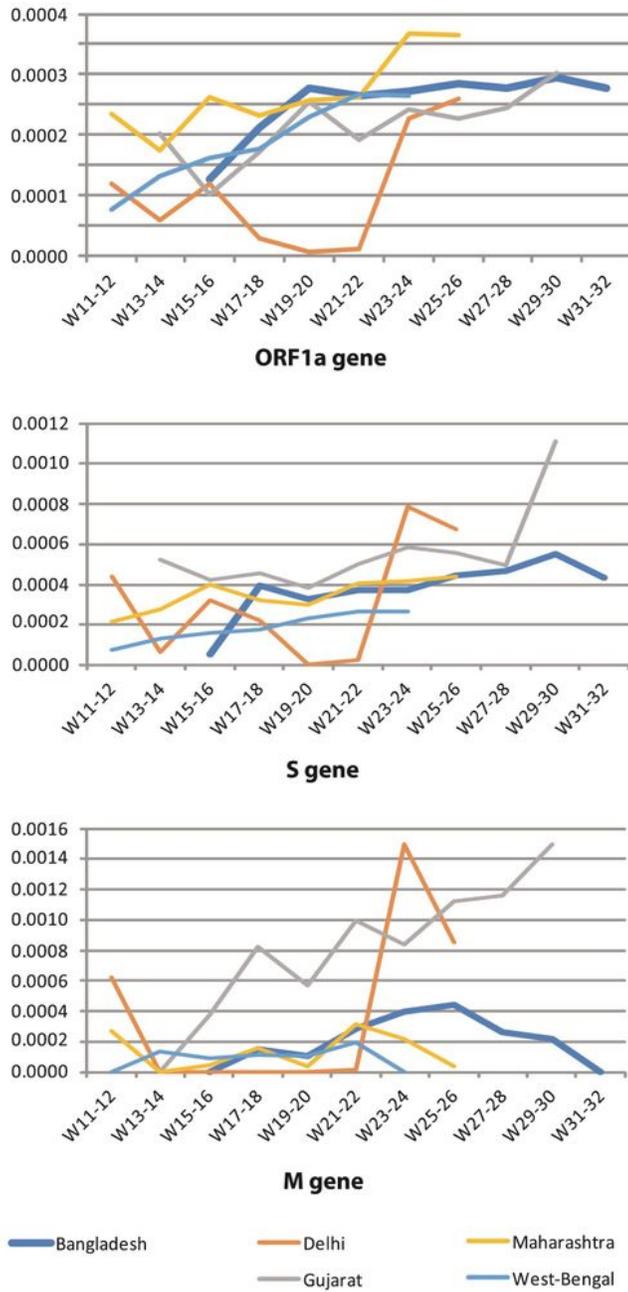


Figure 2

Changes in mean p-distance of the sequence from the reference strain. Y axis represents mean p-distance of three longer genes (ORF1a, S, and M) from the reference strain (Wuhan-Hu-1) throughout two weeks' time-points (in X axis).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.docx](#)
- [SupplementaryTable2.docx](#)
- [SupplementaryTable3.docx](#)
- [SupplementaryTable4.docx](#)