HAL open science

Genomic epidemiology of early SARS-CoV-2 transmission dynamics in Bangladesh

L. Carnegie, J T Mccrone, L. Du Plessis, M. Hasan, M.Z. Ali, R. Begum, M.Z. Hassan, S. Islam, M.H. Rahman, A.S.M. Uddin, et al.

To cite this version:

L. Carnegie, J T Mccrone, L. Du Plessis, M. Hasan, M.Z. Ali, et al.. Genomic epidemiology of early SARS-CoV-2 transmission dynamics in Bangladesh. Virology Journal, 2024, 21 (1), pp.291. 10.1186/s12985-024-02560-2 . hal-04819498

HAL Id: hal-04819498

https://hal.inrae.fr/hal-04819498v1

Submitted on 4 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Carnegie *et al. Virology Journal* (2024) 21:291
https://doi.org/10.1186/s12985-024-02560-2

Virology Journal

RESEARCH Open Access

Genomic epidemiology of early SARS-CoV-2 transmission dynamics in Bangladesh

L. Carnegie[1*†], J. T. McCrone[2,3], L. du Plessis[4,5], M. Hasan[6], M.Z. Ali[6], R. Begum[6], M.Z. Hassan[6], S. Islam[6,7], M.H. Rahman[6], A.S.M. Uddin[6], M.S. Sarker[6], T. Das[8,9], M. Hossain[10,11], M. Khan[12], M.H. Razu[12], A. Akram[13], S. Arina[13], E. Hoque[13], M.M.A. Molla[13], T. Nafisaa[13], P. Angra[14], A. Rambaut[2], S.T. Pullan[15], K.L. Osman[15], M.A. Hoque[8], P. Biswas[8], M.S. Flora[18], J. Raghwani[1], G. Fournié[1,16,17], M.A. Samad[6*†] and S.C. Hill[1*†]

Abstract

Background Genomic epidemiology has helped reconstruct the global and regional movement of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). However, there is still a lack of understanding of SARS-CoV-2 spread in some of the world's least developed countries (LDCs).

Methods To begin to address this disparity, we studied the transmission dynamics of the virus in Bangladesh during the country's first COVID-19 wave by analysing case reports and whole-genome sequences from all eight divisions of the country.

Results We detected > 50 virus introductions to the country during the period, including during a period of national lockdown. Additionally, through discrete phylogeographic analyses, we identified that geographical distance and population -density and/or -size influenced virus spatial dispersal in Bangladesh.

Conclusions Overall, this study expands our knowledge of SARS-CoV-2 genomic epidemiology in Bangladesh, shedding light on crucial transmission characteristics within the country, while also acknowledging resemblances and differences to patterns observed in other nations.

Keywords SARS-CoV-2, Phylodynamics, Bangladesh, Genomic epidemiology

Background

Viral genomic analyses of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have provided insight into viral spread on both local and global scales and helped to inform infection control measures [1–6]. For instance, genomic analysis has been used to identify that Brazil shifted from being a net importer to an exporter of SARS-CoV-2 lineages over the course of 2020 [5], and to reconstruct patterns of within-country viral movements in France during the country's first two Coronavirus disease 2019 (COVID-19) pandemic waves [7]. Using viral genomic data allows for epidemiological insight that cannot be captured from case data alone,

[†]L. Carnegie, M.A. Samad and S.C. Hill contributed equally to this work.

S.C.H and M.S are joint last authors. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention.

*Correspondence:
L. Carnegie
lcarnegie20@rvc.ac.uk
M.A. Samad
msamad@blri.gov.bd
S.C. Hill
sahill@rvc.ac.uk

Full list of author information is available at the end of the article

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

including the inference of transmission dynamics prior to the first recorded sample or during periods with limited disease surveillance.

Despite often having high age-specific infection fatality rates (IFRs) [8], limited genomic studies of SARS-CoV-2 transmission dynamics have been conducted in the world's least developed countries (LDCs) [8–10]. LDCs are low/middle-income countries (LMICs) that the United Nations (UN) classifies as possessing particularly high levels of poverty, human resource weakness, and economic vulnerability [10, 11]. Instead, most studies focus on findings from high income countries (HICs), particularly in Western Europe and North America, likely reflecting a greater ability to invest in both sequence generation and research efforts [8, 9, 12]. This is problematic as findings from HICs may not be readily applicable to LDCs. LDCs often face unique challenges in controlling the virus, including limited public health resources, weak healthcare systems, and large sections of their populations highly susceptible to external economic pressures [8, 9, 13]. The indiscriminate application of control measures that are effective in high-income countries (HICs) may lead to unintended consequences in LDCs, particularly in settings where enforcement is challenging [13]. For example, stay-at-home orders have at times led to mass economic migration from cities, resulting in virus exports to larger geographic areas [14]. Identifying factors that influence transmission dynamics in LDCs allows for better tailoring of control measures to local epidemiological contexts, thereby increasing their chances of success [13, 15].

The limited number of genomic epidemiological analyses in LDCs is exemplified in Bangladesh, where, despite some insights, many aspects of SARS-CoV-2 spread remain poorly understood. Cowley et al., (2021) compared observed mobility pattern data with time-scaled viral phylogenies to suggest that mass migration out of Dhaka, the capital of Bangladesh, may have driven early viral spread across the country [14]. However, in contrast to many HICs, the potential association between spatial and demographic factors and viral dispersal patterns has not been quantified [14]. Likewise, while studies reported that SARS-CoV-2 may have appeared in Bangladesh weeks before the first case report [14] and was likely introduced multiple times from several countries [16–18], the number of viral lineage imports to and exports from Bangladesh have not been estimated.

Bangladesh is the most densely populated country in the world that is not a city-state (165 million inhabitants living at over 1000 people/km^2), and one of several LDCs in South Asia [14, 19–23]. The first case of SARS-CoV-2 in Bangladesh was confirmed on March 8th, 2020 [19, 24]. In response to rising cases, a nationwide lockdown (known as a 'National General Holiday') was introduced on the 26th of March and lasted until the 30th of May. Restrictions and screening procedures were implemented at educational institutions, workplaces and international airports [23, 25]. Infections nevertheless increased steeply in April, with nearly 313,000 confirmed cases reported by 31st August 2020 (a total of 4,281 deaths) [18]. Following the end of the national lockdown, a subsequent zone-coded lockdown was used whereby movement within and between certain high-incidence areas was restricted [23, 25–27].

This study aims to explore the spread of SARS-CoV-2 into and within Bangladesh during the first wave of the COVID-19 pandemic. We generated 175 SARS-CoV-2 whole-genome sequences from cases confirmed by molecular testing. These were combined with 140 additional sequences corresponding to confirmed cases occurring in Bangladesh during the study period to inform a phylodynamic analysis. We assessed the frequency and size of SARS-CoV-2 lineage imports to Bangladesh. We also identified drivers of viral dispersal among regions within the country, using a subset of 194 sequences from the countries' two largest lineages.

Methods

Sample collection

Samples were chosen for sequencing from residual anonymised diagnostic samples collected from patients attending hospitals or health institutes in different districts between 26th April −31st August 2020 that were received by the National Reference Laboratory for Avian Influenza, Bangladesh Livestock Research Institute (BLRI) and that had tested positive by Real-time reverse transcription polymerase chain reaction (rRT-PCR). There was no patient or public involvement in the design of this study. Samples were nasopharyngeal or oropharyngeal swabs or both and were collected along the clinical criteria according to the National Guidelines on Clinical Management of Coronavirus Disease 2019 (COVID-19) published by the Directorate General of Health Services (DGHS), Bangladesh. A structured data collection questionnaire (Figure S1) was collected alongside all samples, detailing each patient's geographic origin, clinical symptoms, travel history, and probable viral source. rRT-PCR for SARS-CoV-2 detection was performed using fluorescent probes and the result was considered positive when the cycle threshold (Ct) values of both the ORF1ab and N genes were <32 and <35, as measured by ROX [28] and FAM [29] dyes [30–32]. We attempted to choose samples proportionally to confirmed cases per district (see Figure S2 for a comparison of sequences to confirmed cases per district), whilst also still including a high percentage of samples from likely under-sampled regions. We preferentially selected

samples for which information on travel history (including confirmation of absence of travel) was present.

Sequencing

We sequenced 175 whole SARS-CoV-2 genomes from samples using a Nanopore GridION and the nCoV-2019 sequencing protocol v3 (LoCost) V.3 incorporating the ARTIC multiplex-PCR sequencing pipeline (as described in [33]). We combined newly generated sequences with 140 Bangladeshi sequences accessed (July 2022) via the GISAID [34] EpiCOV database (www.gisaid.org) collected from patients during the same sampling period. These included: (i) sequences already used in published papers ($n=139$) and (ii) unpublished sequences where submitting laboratories granted us permission for their inclusion ($n=1$). All SARS-CoV-2 sequences were aligned using Llama (Local Lineage and Monophyly Assessment) [35] v0.1 with default parameters. Using AliView [36] v1.28 we confirmed that there were no sequences from the alignment that either were duplicated, short (<70% of the total equence length), or indicative of containing sequencing or assembly errors.

Epidemiological data

We obtained COVID-19 confirmed case data between 17th April −31st August 2020 from the Institute of Epidemiology Disease Control and Research (IEDCR) "COVID-19 Dynamic Dashboard for Bangladesh" [37]. Aggregated daily case counts were available for each of Bangladesh's eight divisions (Chattogram, Dhaka, Sylhet, Khulna, Barishal, Rajshahi, Rangpur, Mymensingh). For the national-level data, days in which data was missing from one or more divisions were excluded ($n=10$) from analyses.

Epidemiological analyses

We applied methods implemented in EpiFilter [38] to estimate the median and 95% equal-tailed Bayesian credible intervals of the effective reproduction number at time *t*, (R(t)), in each division between 17th April 2020 and 31st August 2020. As detailed in [38], this method reduces the mean squared error in R(t) inference by applying optimal recursive smoothing. We used the SARS-CoV-2 serial interval distribution detailed in Ferguson et al. (2020) as the generation time distribution [39]. We also use a weekly averaging filter to minimise the impacts of inconsistent reporting on R(t) inference.

Transmission lineage analyses

We first assembled a global alignment by sampling SARS-CoV-2 genome sequences evenly for each week and country (oldest sequence: 2019-12-24, newest sequence: 2020-08-31). Initially, sequence data was downloaded from GISAID (https://www.gisaid.org) EpiCov database as part of the COG-UK analytical pipeline (https://genomebiology.biomedcentral.com/articles/https://doi.org/10.1186/s13059-021-02395-y) on the 21st April 2021. Only samples with >=90% coverage, a successful pangolin lineage assignment, and a sample date prior to September 1st, 2020, were included in downstream analysis ($n=207,471$). 670 samples from Bangladesh were then removed from the data set ($n=206,801$ remaining) which was then downsampled to 5,000 background samples by hierarchical subsampling so that, as much as possible, samples would be evenly distributed through time, then continent, then country. The 670 Bangladeshi samples from GISAID were added to this background along with the 175 newly generated Bangladeshi sequences with coverage>=90%. Next, temporal outliers, defined as samples that fell beyond 4 interquartile ranges of the expected divergence given were identified and removed ($n=41$, all from the set of 5,000 sequence background samples) using TreeTime [40] and a fixed clock rate of 0.00075 substitutions/site/year. The remaining 5,804 samples were used in downstream analysis. We then used ThorneyBEAST (https://beast.community/thorney_beast) to estimate a posterior molecular clock tree using a pipeline similar to that of du Plessis et al., (2021) [41], Raghwani et al., (2022) [4] and Gutierrez et al., (2021) [42]. This employed a strict molecular clock model [43] (0.00075 substitutions/site/year), and a Skygrid coalescent tree prior [44]. For the starting tree, we used a maximum-likelihood tree produced with IQ-TREE [45] under a Jukes-Cantor substitution model, using Wuhan/WH04/2020 as an outgroup. During the Markov chain Monte Carlo (MCMC), we constrained our tree search to sample only node height and resolutions of polytomies present in the starting tree. We completed 5 chains of 200 million steps with sampling every 1.8 million steps, removing the first 10% of steps as burn-in. The resulting sample of 500 trees from the posterior were used to generate a maximum clade credibility (MCC) tree. We confirmed that the Effective Sample Size (ESS) of all parameters was >200 using Tracer [46] v1.7.1.

We then conducted a discrete trait analysis (DTA) with these 500 empirical trees in BEAST [47] to estimate ancestral node locations (Bangladesh or non-Bangladesh). We used a robust counting approach [48] to evaluate the expected number of location state transitions into and out of Bangladesh across a posterior sample of 2000 trees.

Finally, we used the location-annotated MCC tree to identify transmission lineages in Bangladesh, following previously described methods [41]. Briefly, nodes with a posterior probability >0.5 inferred to be located in Bangladesh are considered "Bangladeshi nodes". A depth-first search initiated at a random Bangladeshi node is run until complete or until a non-Bangladeshi node is

encountered, with encountered nodes labelled as belonging to the same transmission lineage. A new starting node is chosen until every Bangladeshi node is visited, and hence all Bangladeshi tips have been assigned a lineage or classified as a singleton. For each Bangladeshi lineage, we estimated time to the most recent common ancestor (TMRCA).

Discrete trait analysis with generalized linear model

To investigate how viral dispersal was influenced by spatial and demographic factors in Bangladesh we focused on the two largest transmission lineages identified in the global dataset (Material and Methods section "Transmission lineage analyses") Only those whole-genome sequences with known sampling districts were used (n=142/307 for lineage 2 and n=52/97 for lineage 8, remaining total with sampling districts=194) (Tables S1-S3). To facilitate effective discrete phylogeography analyses, we reduced the original 28 districts (second administrative level), covering six out of the eight Bangladeshi divisions (first administrative level) into a smaller set of discrete locations. In brief, we employed k-means clustering based on Euclidean distance, as implemented in the '*stats*' package in R [49] v4.1.2, to group these sequences into eight geographic regions, each representing a grouping of several districts. All sequences were assigned to their respective district centroid coordinates for this purpose. We assembled data on chosen predictors for each of these eight geographical groups of districts: population size and density, the ratio between the number of cases and population size, and geographic distance (ellipsoid distance between grouping centroids) (see Tables S4-S5 for details). Whilst we did not expect a strong temporal signal given the low mutation rate of SARS-CoV-2 and the short timescale of our study (e.g [14, 50]), we used TempEst [51] v1.5.3 to check that the temporal signal in each lineage alignment was not incompatible with using fixed substitution rates based on the literature and to check for outlier sequences (Figure S3). We used a DTA model implemented in BEAST [47] v1.10.4 to reconstruct virus lineage movements between the eight pre-defined geographic groupings. A strict molecular clock model with a fixed rate (0.001 substitutions/site/year) and an SRD06 substitutional model [52] was shared between both transmission lineages to improve convergence, with a separate exponential growth coalescent tree prior that allows for different timings of lineage growth, and default BEAST [47] v1.10.4 priors for the other parameters. We used a faster clock rate (0.001 substitutions/site/year) in these shorter timescale phylogeographic analyses, compared to the previous global transmission lineages analyses (0.00075 substitutions/site/year), to account for the existence of slightly deleterious mutations circulating in the Bangladesh transmission lineages that had not been removed by purifying selection [53]. We used separate generalised-linear models (GLMs) to determine which covariate(s) (Table S4-S5) best predicted the frequency of viral lineage movements between locations for each lineage. The migration rates in the DTA model are parameterised as a GLM using covariates as coefficients of the GLM. DTA-GLM analyses that included either population density or population size predictors were run separately because of the multicollinearity between these predictors. For each analysis, we executed two chains of 150 million steps logged every 15,000 steps, removing the first 10% of steps as burn-in. We used Tracer [46] v1.7.1 to confirm that the ESS was >200 for all parameters. The support of the Bayes Factor (BF) for transmission between discrete traits was interpreted as described previously in [54], with BF>100 'very strong support', and BF>10 indicating 'strong support'.

Results

Epidemiological findings

COVID-19 confirmed cases increased towards the end of the national lockdown and continued rising during the subsequent zone-coded lockdown, peaking at the start of July 2020 (Fig. 1). Reported case numbers dropped sharply during Eid al-Adha, which is one of the most important Muslim festivals and marked by a period of public holiday in Bangladesh [55] (31st July 2020–1st August 2020) (Fig. 1). This is consistent with reduced reporting and testing in other countries in periods surrounding important religious events [56].

We plotted daily COVID-19 confirmed case counts in Bangladesh's eight divisions between 17th April 2020 and 31st August 2020. Case counts in each division were similar to the pattern observed at the country level (Figs. 1 and 2A). It is possible that rising cases could be somewhat driven by increased testing capacity over time [14]. In particular, during the first months of the pandemic, residents in Dhaka, the country's largest and most populated city, had easier access to SARS-CoV-2 testing compared to other regions in Bangladesh [14, 57]. Consequently, the Dhaka division reported the majority (50.5%) of confirmed SARS-CoV-2 cases in the country (Fig. 2A) [14, 57]. Spikes in reported cases during the first wave of the pandemic were observed in some other divisions, e.g., in Rangpur in the start of July 2020 (Fig. 2A).

Division-specific effective reproduction numbers R(t) showed similar temporal trends in most Bangladeshi divisions (Fig. 2B). R(t) typically fell in the first half of the national lockdown, but then peaked above 1 near the middle or end of May, concurrent with both the religious festival of Eid al-Fitr and the end of the national lockdown. R(t) then slowly decreased and fluctuated around 1 by the start of July (Fig. 2B). Division-specific spikes

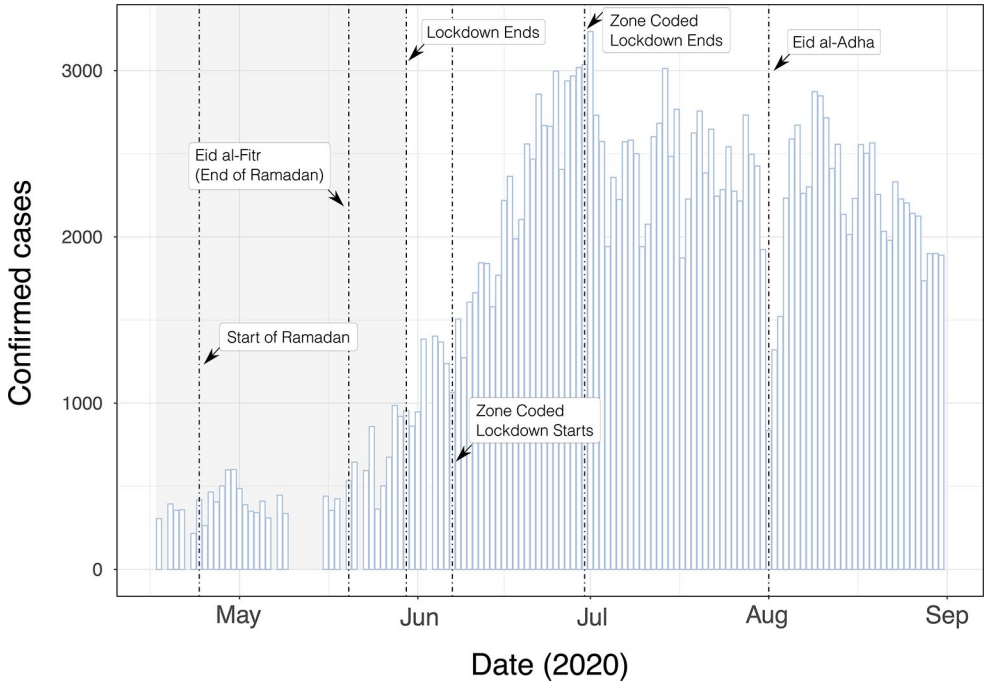

Fig. 1 Daily COVID-19 confirmed case counts in Bangladesh, 2020. Grey highlighted region indicates the timing of a national lockdown. The x-axis ticks correspond to the start of the named month

and troughs in R(t) were also observed, such as a reduction of R(t) below 1 in mid-June in Mymensingh (Fig. 2B).

Transmission lineage findings

We estimated the number of SARS-CoV-2 lineage importations over time to Bangladesh. The estimated number of viral lineage movements into the country ($n=111$, 95% HPD: 99–121) was approximately 4 times higher than the number of viral lineage movements from Bangladesh to other countries globally ($n=26$, 95% HPD: 16–38) (Fig. 3A), indicating that Bangladesh was mainly an importer of virus lineages during the first wave.

We detected 56 distinct Bangladesh transmission lineages (i.e., lineages spreading within Bangladesh, represented by >1 genome), and 51 singletons (singular Bangladesh genomes representing lineage imports originating from non-Bangladesh regions, yet without evidence of autochthonous transmission) (Figure S4). Twelve lineages contained>10 sequences, and two lineages contained>100 genomes. Several lineages ($n=16$, 28.6%), including the two largest lineages, had an estimated TMRCA that fell either prior to or within the following week of the start of the national lockdown on the 26th March (Fig. 3B & S4) [25]. Multiple lineages, however, had TMRCAs overlapping with the period of nationwide lockdown (Fig. 3C & S4). Finally, there was a strong negative exponential relationship between the estimated TMRCA and transmission lineage size (Fig. 3C). Larger transmission lineages in Bangladesh were generally associated with longer sampling durations (Fig. 3D).

Drivers of virus spread within Bangladesh

We used DTA models coupled with GLMs to identify significant predictors of virus lineage dissemination among eight geographical groups each containing districts from multiple Bangladeshi divisions (Figs. 4A and 5), using only those sequences from the two largest Bangladesh transmission lineages (Figs. 3 and 5). We fitted two models with either population size or population density as the demographic predictor (covariate values shown in Table S5 & Figure S5). We observed consistent directionality and effect sizes for two population metrics in each of their respective models, but notable differences were found in the BF support and inclusion probabilities. In the model including population density as a predictor (Fig. 4B), there was strong evidence (BF>10, following: [54]) that virus lineages were more likely to move from areas with high population density. We also found very strong support (BF>100) for virus movement between geographically close locations and very strong support (BF>100) towards areas with a high number of confirmed cases per capita. In the model including population size as predictor (Fig. 4C), there was strong support for virus spread occurring more commonly to and from areas with high population sizes, and for virus movement towards locations with a high number of confirmed cases

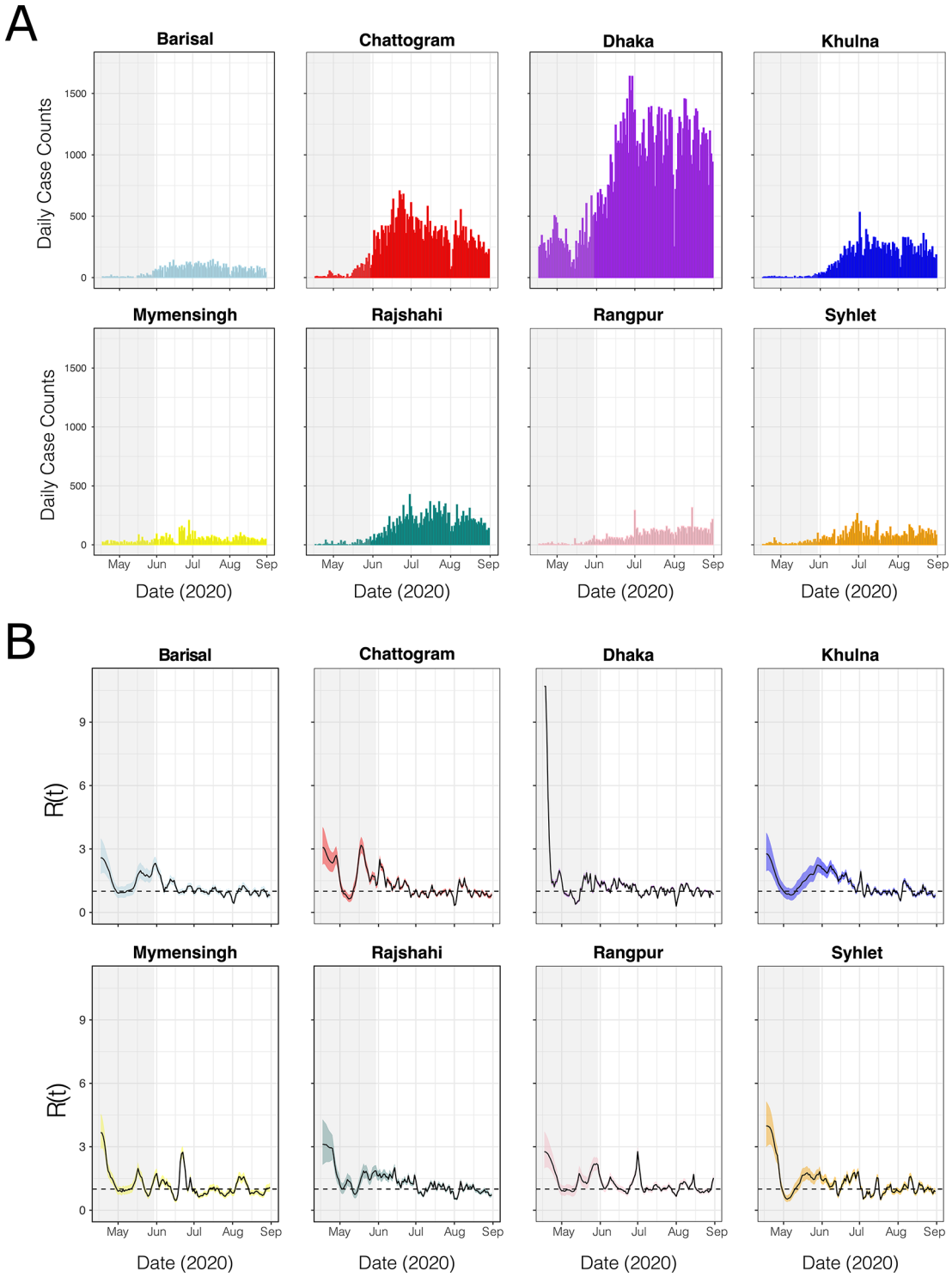

Fig. 2 Grey-highlighted regions indicate a national lockdown. (**A**) Daily COVID-19 confirmed cases in each division. (**B**) Effective reproduction number (R(t)) by division. Median Rt estimates are shown by black lines, and 95% Bayesian credible intervals by coloured ribbons

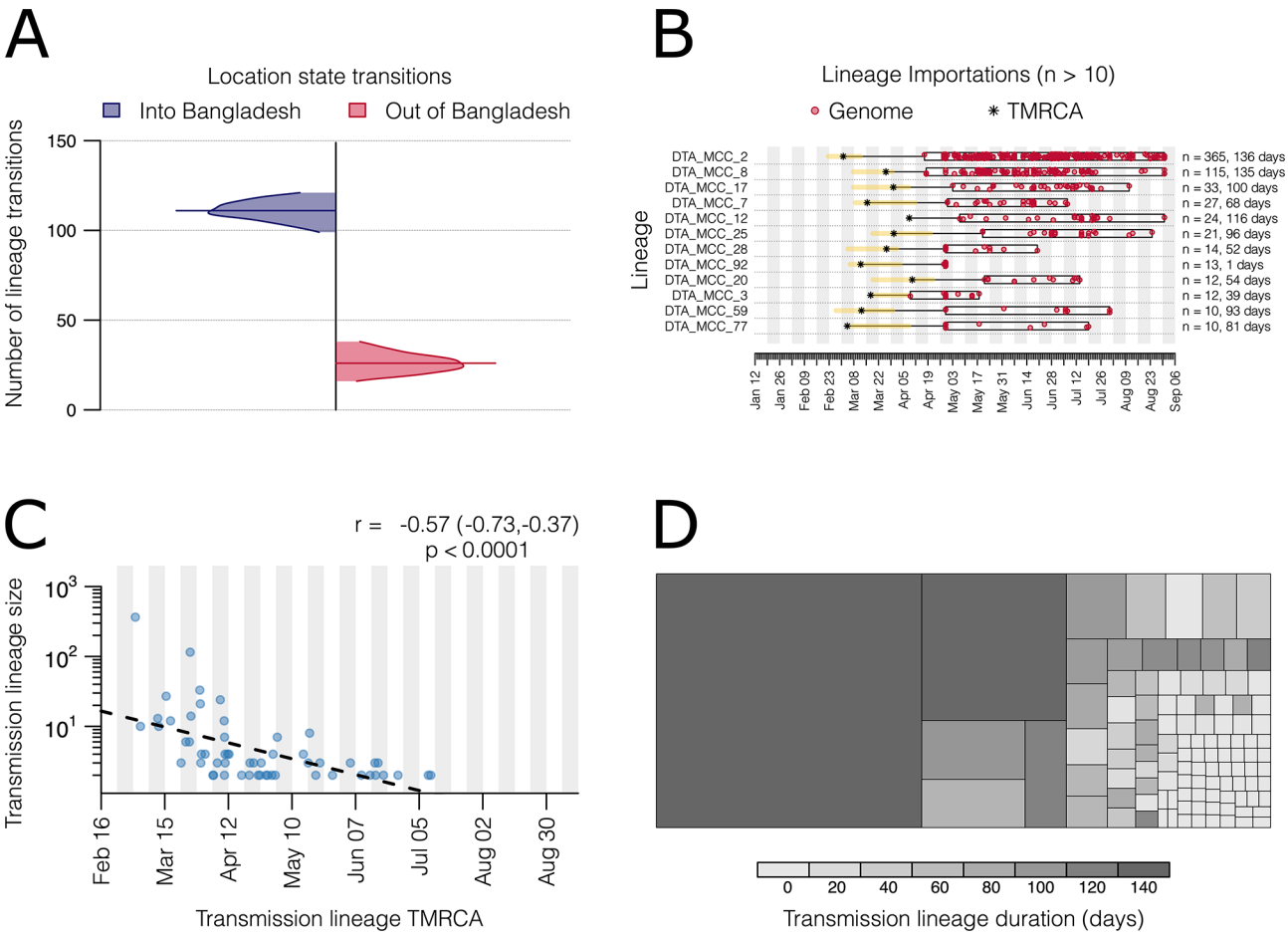

Fig. 3 SARS-CoV-2 transmission lineage characteristics in Bangladesh. (**A**) Number of location state transitions between the phylogenetic traits Bangladesh/Global (imports into Bangladesh = blue, exports from Bangladesh = red), as detected via the robust counting approach. Posterior distributions are truncated at their 95% highest posterior distribution (HPD) interval limits and median estimates are shown using horizontal lines. (**B**) Duration and timing of the largest Bangladesh transmission lineages (> 10 genomes). Each row represents a transmission lineage, and red dots indicate genome sampling times. Boxes and labels on the right axis show the sampling duration (see Figure S3 for more details on sampling duration per lineage), and number of sampled genomes (n). Asterisks show the median estimated time to most recent common ancestor (TMRCA) for each lineage, with the 95% HPD as a yellow bar (**C**) Relationship between transmission lineage size and TMRCA, with a dashed line indicating the slope of a linear regression. The Pearson correlation coefficient, 95% confidence interval, and p-value are shown. (**D**) Partition of Bangladesh genomes into cells representing transmission lineages and singletons, each coloured by estimated duration (time between the lineage's oldest and most recent genomes). Cell size is proportional to lineage size

per capita. However, distance was not identified as a significant predictor (BF < 1) (See Fig. 4C).

Discussion

Genomic epidemiology has played a crucial role in understanding the global and regional movement of SARS-CoV-2 [1–6]. However, research efforts have disproportionately neglected studying virus spread within the world's least developed countries. To contribute to addressing this disparity, we use a combination of epidemiological and phylodynamic analyses to investigate the transmission dynamics of SARS-CoV-2 in Bangladesh during the country's initial wave of COVID-19. We determined that R(t) showed a declining trend during the start of the national lockdown but rose slightly towards the end of the national lockdown in late May 2020, which was perhaps driven by activity surrounding Eid al-Fitr. Subsequently, R(t) slowly stabilised and fluctuated around ~ 1 in all divisions. Our study revealed that cases in Bangladesh were initiated by more than 50 instances of virus introduction into the country, with multiple lineages maintained even during the period of national lockdown. Finally, we identified using discrete phylogeographic analyses that patterns of virus dispersal may have been shaped by population density and/or size, as well as geographical proximity, although the latter was not consistently important across all analyses.

The detection of over 50 virus introductions during Bangladesh's initial COVID-19 wave, and of multiple persistent transmission lineages during the lockdown, contrasts with studies from Hong Kong [58] and New Zealand [59], where virus cases were linked to repeated

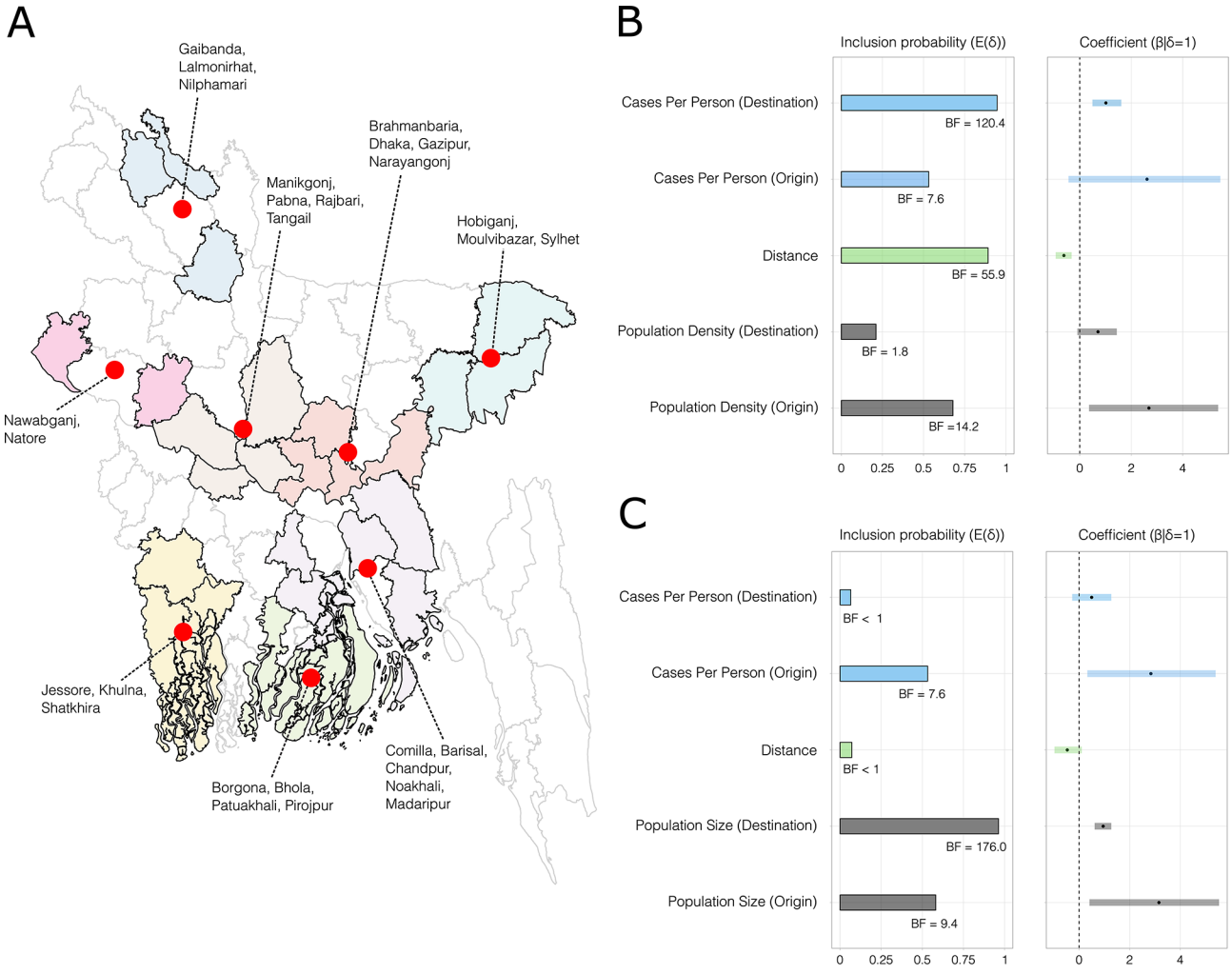

Fig. 4 Factors associated with SARS-CoV-2 spread. **A**) Map of Bangladesh showing the eight geographical groups of districts used as discrete traits in the DTA-GLM. Region centroids are marked by red dots. **B** and **C**) Predictors of SARS-CoV-2 spread based on models with either population density (**B**) or population size (**C**) included as predictors. Bar and line colours indicate different covariates, with origin and destination predictor of a covariate given the same colour within each plot. Inclusion probability is the posterior expectation that the indicator variable is associated with each predictor E(δ) and suggests that the predictor is associated with different rates of viral diffusion. Bayes Factor (BF) support values for each covariate are indicated by black text annotations. The coefficient (β|δ = 1) represents the contribution of each predictor on a log scale when the predictor is included in the model, with the 95% credible interval of the GLM coefficients (β) represented by horizontal lines

individual introductions rather than persistently present lineages. Both our study and the Hong Kong study were based on similar pipelines adapted from du Plessis et al. (2021) [41], and therefore the observed difference is likely not caused by different study designs. Instead, the distinction likely reflects a greater inability to adequately maintain social distancing in Bangladesh's high-density urban areas and the relatively greater capacity of these other countries to control local transmission through measures such as community surveillance and contact tracing [12, 44]. Additionally, New Zealand and Hong Kong adopted stringent inbound travel regulations and testing measures [58, 59]. In contrast, Bangladesh had limited thermal scanners at Dhaka international airports, and reported low quarantine compliance [19, 60]. These differences may have played a further role in the observed variations in SARS-CoV-2 lineage dynamics between locations.

In congruence with previous studies in South Africa [6] and Italy [61], we found that the estimated number of virus importations into the country considerably outweighed exports. The return of a large number of Bangladeshi nationals to the country during the early stages of the pandemic may have allowed for extensive lineage importation to the country from a wide range of geographical locations [14, 16–18, 62]. Notably, over 10 million Bangladeshi citizens live abroad, and the number of migrant labourers returning to the country was 8 times higher in 2020 than in previous years [18, 63, 64]. Introductions may have been seeded via international

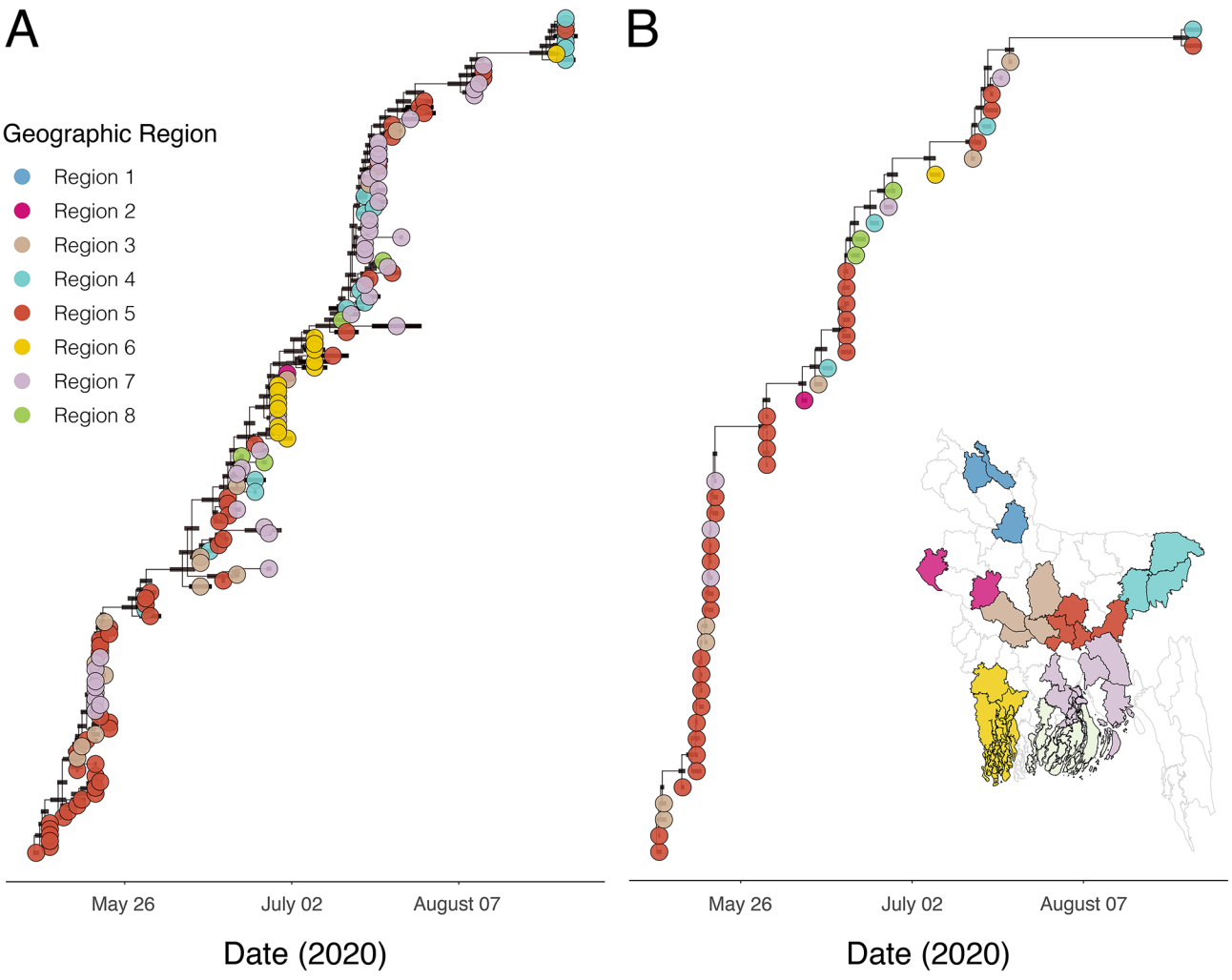

Fig. 5 MCC phylogenies of the two largest lineages detected. (**A**) Lineage 2, and (**B**) Lineage 8. Branch lengths represent time, as shown on the axis. Tips are coloured by the sampling geographic region used in the DTA-GLM analyses, as shown in the inset map. Black bars indicate the 95% highest posterior density (HPD) interval for node ages

airports [14, 16–18, 60], as well as through cross-border human mobility with neighbouring countries (e.g., India [65–67]). The latter has been an important pathway for the introduction of SARS-CoV-2 lineages to other LDCs (e.g., Burkina Faso [68] and Rwanda [69]). However, as the number and origin of introductions and exports estimated from genomic data are dependent on sample size [41], our analyses may be biased towards capturing imports rather than exports, given relatively less intense sampling in Bangladesh compared to many other regions globally.

Viral lineages moved more frequently from locations with higher to locations with lower population density or total population size. This finding may reflect the spread of viral lineages across the country during the early stages of the pandemic being driven by mass migration out of major cities/urban hotspots with relatively higher case rates to rural areas, following the announcement or extension of the national lockdown, as previously suggested by several publications that reference news reports (e.g [19, 70]) and a comparative analyses of mobility patterns with genomic data [14]. The considerable number of commuters from peripheral rural locations into urban areas may also have contributed to virus spread following the ease of the national lockdown [14, 71–73]. This pattern of wide-scale movement from urban areas driving early spread of the virus has previously been identified using genomic methods in both HICs (e.g., France [7]) and regions of lower-middle-income countries (LMICs) (e.g., Gujarat, India [4]). As such, in conjunction with this previous research, our study's findings may further support the notion that population size/density and mass human movements were generally important drivers of within-country spread during this stage of the pandemic, regardless of country wealth [74].

In one of our discrete phylogeographic analyses, we identified that viral movement occurred preferentially between more proximate geographic groupings of districts in Bangladesh. This suggests that certain lockdown measures (e.g., travel advisories, railway closures, domestic flight suspensions [23, 25–27]), likely had a more substantial impact on restricting long-distance human mobility and virus spread, compared to movement and viral dispersal between closer or neighbouring divisions. Distance has also been identified as a predictor of SARS-CoV-2 spread in several other countries, including Israel and Brazil [75, 76]. That said, the presence of strong evidence (using BF support) for distance depended here on whether human population density or size was considered, with only strong evidence detected when the population density predictor was also included in the model. This could indicate that distance had a relatively weak effect on viral transmission dispersal between regions in the country, perhaps because Bangladesh is relatively small and its population is highly mobile [72].

Our study has several limitations. Firstly, only a very small fraction of the total confirmed SARS-CoV-2 cases could be sequenced due to resource constraints. Greater sampling and sequencing of cases would have allowed us to analyse the virus dispersal patterns in Bangladesh at higher resolution. Second, the limited genomic surveillance in certain locations outside of Dhaka where many cases were confirmed, particularly the absence of sequences from Chattogram, the country's second-largest city, indicates that our analyses may have underestimated the extent and frequency of virus movements within Bangladesh. As such, the outputs of our DTA-GLM models may have not fully captured the strength or impact that certain demographic predictors had on within-country viral dispersal [14]. We attempted to mitigate this limitation by choosing samples for sequencing roughly proportionally to the confirmed cases per Bangladeshi district while retaining most samples from any particularly under sampled locations. However, for several regions, very few samples were available and case reporting is likely to be inherently spatially biased. Third, due to the short time-period of our study, we were unable to investigate changes in the drivers of viral spread during the study period. The dynamics of transmission and the sources of infection may have altered over time, especially in response to mobility restrictions and interventions such as border closures. Furthermore, the geographic origin of each sequence was determined based on individuals voluntarily disclosing their present place of residence, which can be prone to inaccuracies. Likewise, missing location metadata meant we had to reduce the number of sequences from each lineage to only sequences with known division when performing the DTA-GLM phylogeographic analyses, which may have reduced the possible statistical power of the analyses.

Conclusions

In conclusion, this study on SARS-CoV-2 genomic epidemiology during the first wave in Bangladesh uncovers distinct and common virus transmission patterns in comparison to other countries of various income levels. These findings highlight the significance of genomic epidemiological analyses in resource-constrained regions like Bangladesh in potentially helping to inform the design of more specific and accurate evidence-based interventions aimed at reducing the import and within-country spread of viral outbreaks. Addressing the limitations in genomic surveillance in Bangladesh is crucial for improving the capacity to respond to future viral outbreaks.

Abbreviations

BF	Bayes Factor
COVID-19	Coronavirus disease 2019
Ct	Cycle threshold
DTA	Discrete trait analysis
ESS	Effective Sample Size
GLMs	Generalised-linear models
HICs	High income countries
LDCs	Least developed countries
LMICs	Low/middle-income countries
MCC tree	Maximum clade credibility tree
MCMC	Markov chain Monte Carlo
rRT-PCR	Real-time reverse transcription polymerase chain reaction
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
TMRCA	Time to the most recent common ancestor
UN	United Nations

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12985-024-02560-2.

Supplementary Material 1

Acknowledgements

We thank the Child Health Research Foundation (CHRF), Bangladesh, for allowing us to use their data on the GISAID EpiCoV database (accession numbers in Table S7-S8). We also acknowledge the use of additional sequences from the GISAID EpiCoV database for the global dataset used in our lineage analyses and provide accession numbers and acknowledgements in the supplementary electronic material (Table S7-S8). We thank the Directorate General of Health Services (DGHS) and Institute of Epidemiology Disease Control and Research (IEDCR) Bangladesh, DGHS, Ministry of Health and Family Welfare for sharing COVID-19 samples and data.

Author contributions

L.C., S.C.H, G.F. and J.R. conceived the phylodynamic and epidemiological analysis. L.C. wrote the manuscript and performed the phylodynamic and epidemiological analysis. S.C.H, J.R. and G.F. assisted with phylodynamic and epidemiological analysis. JT.M. and L.D.P. performed the viral lineage analyses. S.C.H., G.F. and J.R. edited the manuscript and supervised the work. M.A.S, M.H., M.S.S., and S.I. contributed to the collection of epidemiological data, laboratory testing, and generation of sequences and helped with interpreting the findings. All other authors contributed to fieldwork or laboratory work design or implementation. All authors gave final approval for publication.

Funding
L.C. is supported by a Biotechnology and Biological Sciences Research Council (BBSRC) London Interdisciplinary Biosciences Consortium (LIDo) studentship [grant number BB/T008709/1]. S.C.H. is supported by a Sir Henry Wellcome Postdoctoral Fellowship from the Wellcome Trust [220414/Z/20/Z] (https://wellcome.org/). J.R., M.H. and G.F. are supported by the UKRI GCRF One Health Poultry Hub (Grant No. BB/S011269/1), one of twelve interdisciplinary research hubs funded under the UK government's Grand Challenge Research Fund Interdisciplinary Research Hub initiative. G.F. was supported by the French National Research Agency and the French Ministry of Higher Education and Research. M.A.S. is supported by Zoonosis and Transboundary Animal Diseases Prevention and Control project [Grant No.223038200], MoFL, Bangladesh. S.I., R.B., and M.S.S. are supported by the 'Combating the threats of antimicrobial resistance and zoonotic diseases to achieve the GHSA in Bangladesh' [Grant no. NU2GGH002077].

Data availability
The datasets and BEAST XML files supporting the conclusions of this article are available in the 'SARS_CoV_2_Bangladesh"repository, [https://github.com/lorcancarnegie/SARS_CoV_2_Bangladesh]. The dataset for the newly reported genetic sequence data supporting the conclusions of this article are available in GISAID EpiFlu database (https://gisaid.org) under the accession numbers included within the additional files of this article [Table S6].

Declarations

Ethics approval and consent to participate
Patient consent for publication: Not applicable. The study was carried out under the ethical approval from the ethical approval committee of Bangladesh Livestock Research Institute, Bangladesh with the reference number BLRI/EA/2020102/2022, and with informed consent from tested individuals.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Pathobiology and Population Sciences, Royal Veterinary College (RVC), Hatfield, Hertfordshire, UK
[2]Institute of Ecology and Evolution, University of Edinburgh, King's Buildings, Edinburgh, UK
[3]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[4]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland
[5]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[6]Bangladesh Livestock Research Institute (BLRI), Savar, Dhaka, Bangladesh
[7]Global Change Center, Virginia Tech, Blacksburg, VA, USA
[8]Chattogram Veterinary and Animal Sciences University (CVASU), Khulshi, Chattogram, Bangladesh
[9]School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia
[10]NSU Genome Research Institute (NGRI), North South University, Bashundhara, Dhaka, Bangladesh
[11]Department of Biochemistry and Microbiology, North South University, Bashundhara, Dhaka, Bangladesh
[12]Bangladesh Reference Institute for Chemical Measurements (BRiCM), Dhanmondi, Dhaka, Bangladesh
[13]National Institute of Laboratory Medicine and Referral Centre (NILMRC), Agargoan, Dhaka, Bangladesh
[14]Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA
[15]United Kingdom Health Security Agency (UKHSA), Porton Down, Salisbury, UK
[16]Université de Lyon, INRAE, VetAgro Sup, UMR EPIA, Marcy l'Etoile, France
[17]Université Clermont Auvergne, INRAE, VetAgro Sup, UMR EPIA, Saint Genes Champanelle, France
[18]National Institute of Preventive and Social Medicine (NIPSOM), Ministry of Health and Family Welfare, Dhaka, Bangladesh

Received: 12 March 2024 / Accepted: 26 October 2024
Published online: 13 November 2024

References

1. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, et al. Genomic epidemiology of SARS-CoV-2 in Guangdong Province. China Cell. 2020;181:997–e10039. https://doi.org/10.1016/j.cell.2020.04.023.
2. Louis du P T, MJ EZA, Verity H, Christopher R, Bernardo G et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science (1979) 2021;371:708–12. https://doi.org/10.1126/science.abf2946
3. Featherstone LA, Di Giallonardo F, Holmes EC, Vaughan TG, Duchêne S. Infectious disease phylodynamics with occurrence data. Methods Ecol Evol. 2021;12:1498–507. https://doi.org/10.1111/2041-210X.13620.
4. Raghwani J, du Plessis L, McCrone JT, Hill SC, Parag KV, Thézé J, et al. Genomic epidemiology of early SARS-CoV-2 Transmission dynamics, Gujarat, India. Emerg Infect Dis. 2022;28:751–8. https://doi.org/10.3201/eid2804.212053.
5. Giovanetti M, Slavov SN, Fonseca V, Wilkinson E, Tegally H, Patané JSL, et al. Genomic epidemiology of the SARS-CoV-2 epidemic in Brazil. Nat Microbiol. 2022;7:1490–500. https://doi.org/10.1038/s41564-022-01191-z.
6. Tegally H, Wilkinson E, Lessells RJ, Giandhari J, Pillay S, Msomi N, et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. Nat Med. 2021;27:440–6. https://doi.org/10.1038/s41591-021-01255-3.
7. Coppée R, Blanquart F, Jary A, Leducq V, Ferré VM, Franco Yusti AM et al. Phylodynamics of SARS-CoV-2 transmissions in France, Europe and the world during 2020 2022. https://doi.org/10.1101/2022.08.10.22278636
8. Levin AT, Owusu-Boaitey N, Pugh S, Fosdick BK, Zwi AB, Malani A, et al. Assessing the burden of COVID-19 in developing countries: systematic review, meta-analysis and public policy implications. BMJ Glob Health. 2022;7:e008477. https://doi.org/10.1136/bmjgh-2022-008477.
9. Usuzaki T, Chiba S, Shimoyama M, Ishikuro M, Obara T. A disparity in the number of studies related to COVID-19 and SARS-CoV-2 between low- and middle-income countries and high-income countries. Int Health. 2021;13:379–81. https://doi.org/10.1093/inthealth/ihaa088.
10. Nations UUN. accessed February 7, List of LDCs n.d. https://www.un.org/ohrlls/content/list-ldcs (2024).
11. Nations U. accessed February 7, UN: LDC Identification Criteria & Indicators n.d. https://www.un.org/development/desa/dpad/least-developed-country-category/ldc-criteria.html (2024).
12. Brito A, Semenova E, Dudas G, Hassler G, Kalinich C, Kraemer M, et al. Global disparities in SARS-CoV-2 genomic surveillance. Nat Commun. 2022;13:7003. https://doi.org/10.1038/s41467-022-33713-y.
13. Hodgins S, Saad A. Will the higher-income country blueprint for COVID-19 work in low- and lower Middle-Income countries? Glob Health Sci Pract. 2020;8:136–43. https://doi.org/10.9745/GHSP-D-20-00217.
14. Cowley LA, Afrad MH, Rahman SIA, Mamun MM, AI, Chin T, Mahmud A, et al. Genomics, social media and mobile phone data enable mapping of SARS-CoV-2 lineages to inform health policy in Bangladesh. Nat Microbiol. 2021;6:1271–8. https://doi.org/10.1038/s41564-021-00955-3.
15. Laxminarayan R, Wahl B, Dudala SR, Gopal K, Mohan BC, Neelima S, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. Science. 2020;370:691–7. https://doi.org/10.1126/science.abd7672.
16. Shishir TA, Naser I, Bin, Faruque SM. In silico comparative genomics of SARS-CoV-2 to determine the source and diversity of the pathogen in Bangladesh. PLoS ONE. 2021;16:e0245584.
17. Sharif N, Dey SK. Phylogenetic and whole genome analysis of first seven SARS-CoV-2 isolates in Bangladesh. Future Virol. 2020;15:735–46. https://doi.org/10.2217/fvl-2020-0201.
18. Parvez MSA, Rahman MM, Morshed MN, Rahman D, Anwar S, Hosen MJ. Genetic analysis of SARS-CoV-2 isolates collected from Bangladesh: insights into the origin, mutational spectrum and possible pathomechanism. Comput Biol Chem. 2021;90:107413. https://doi.org/10.1016/j.compbiolchem.2020.107413.
19. Anwar S, Nasrullah M, Hosen MJ. COVID-19 and Bangladesh: challenges and how to address them. Front Public Health. 2020;8:154. https://doi.org/10.3389/fpubh.2020.00154.
20. Sharif N, Dey SK. Impact of population density and weather on COVID-19 pandemic and SARS-CoV-2 mutation frequency in Bangladesh. Epidemiol Infect. 2021;149:e16. https://doi.org/10.1017/S0950268821000029.
21. Streatfield PK, Karar ZA. Population challenges for Bangladesh in the coming decades. J Health Popul Nutr. 2008;26:261–72.

22. Khan MR, Huq S, Risha AN, Alam SS. High-density population and displacement in Bangladesh. Science (1979) 2021;372:1290–3. https://doi.org/10.1126/science.abi6364
23. Ahammad I, Hossain MU, Rahman A, Chowdhury ZM, Bhattacharjee A, Das KC, et al. Wave-wise comparative genomic study for revealing the complete scenario and dynamic nature of COVID-19 pandemic in Bangladesh. PLoS ONE. 2021;16:e0258019.
24. Saha O, Islam I, Shatadru RN, Rakhi NN, Hossain MS, Rahaman MdM. Temporal landscape of mutational frequencies in SARS-CoV-2 genomes of Bangladesh: possible implications from the ongoing outbreak in Bangladesh. Virus Genes. 2021;57:413–25. https://doi.org/10.1007/s11262-021-01860-x.
25. Islam A, Sayeed MA, Rahman MK, Zamil S, Abedin J, Saha O, et al. Assessment of basic reproduction number (R(0)), spatial and temporal epidemiological determinants, and genetic characterization of SARS-CoV-2 in Bangladesh. Infect Genet Evol. 2021;92:104884. https://doi.org/10.1016/j.meegid.2021.104884.
26. Hossain MA, Jahid MIK, Hossain KMA, Walton LM, Uddin Z, Haque MO, et al. Knowledge, attitudes, and fear of COVID-19 during the Rapid rise period in Bangladesh. PLoS ONE. 2020;15:e0239646. https://doi.org/10.1371/journal.pone.0239646.
27. Shammi M, Bodrud-Doza M, Islam ARMT, Rahman MM. Strategic assessment of COVID-19 pandemic in Bangladesh: comparative lockdown scenario analysis, public perception, and management for sustainability. Environ Dev Sustain. 2021;23:6148–91. https://doi.org/10.1007/s10668-020-00867-y.
28. Cheah ESG, Malkin J, Free RC, Lee S-M, Perera N, Woltmann G, et al. A two-tube combined TaqMan/SYBR Green assay to identify mycobacteria and detect single global lineage-defining polymorphisms in Mycobacterium tuberculosis. J Mol Diagn. 2010;12:250–6. https://doi.org/10.2353/jmoldx.2010.090030.
29. Witham PK, Yamashiro CT, Livak KJ, Batt CA. A PCR-based assay for the detection of Escherichia coli Shiga-like toxin genes in ground beef. Appl Environ Microbiol. 1996;62:1347–53. https://doi.org/10.1128/aem.62.4.1347-1353.1996.
30. Dip SD, Sarkar SL, Setu MAA, Das PK, Pramanik MHA, Alam ASMRU, et al. Evaluation of RT-PCR assays for detection of SARS-CoV-2 variants of concern. Sci Rep. 2023;13:2342. https://doi.org/10.1038/s41598-023-28275-y.
31. Lu Y, Li L, Ren S, Liu X, Zhang L, Li W, et al. Comparison of the diagnostic efficacy between two PCR test kits for SARS-CoV-2 nucleic acid detection. J Clin Lab Anal. 2020;34:e23554. https://doi.org/10.1002/jcla.23554.
32. Wen D, Yang S, Li G, Xuan Q, Guo W, Wu W. Sample-to-answer and routine real-time RT-PCR: a comparison of different platforms for SARS-CoV-2 detection. J Mol Diagn. 2021;23:665–70. https://doi.org/10.1016/j.jmoldx.2021.02.010.
33. Tyson JR, James P, Stoddart D, Sparks N, Wickenhagen A, Hall G et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. BioRxiv. 2020. https://doi.org/10.1101/2020.09.04.283077
34. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 2017;22. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.
35. O'Toole A, Hill V. Llama Software: local lineage and monophyly assessment v0.1 2020.
36. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30:3276–8. https://doi.org/10.1093/bioinformatics/btu531.
37. COVID-19 Dynamic Dashboard for Bangladesh. Institute of Epidemiology Disease Control and Research n.d. http://103.247.238.92/webportal/pages/covid19.php#
38. Parag KV. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. PLoS Comput Biol. 2021;17:e1009347. https://doi.org/10.1371/journal.pcbi.1009347.
39. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. 2020.
40. Sagulenko P, Puller V, Neher RA, TreeTime. Maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4:vex042. https://doi.org/10.1093/ve/vex042.
41. du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science (1979) 2021;371:708–12. https://doi.org/10.1126/science.abf2946
42. Gutierrez B, Márquez S, Prado-Vivar B, Becerra-Wong M, Guadalupe JJ, Candido DDS, et al. Genomic epidemiology of SARS-CoV-2 transmission lineages in Ecuador. Virus Evol. 2021;7:veab051. https://doi.org/10.1093/ve/veab051.
43. Marco AR, Ferreira MAS. Bayesian analysis of elapsed times in continuous-time Markov chains. Can J Stat. 2008;36:355–68.
44. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol Biol Evol. 2013;30:713–24. https://doi.org/10.1093/molbev/mss265.
45. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev/msu300.
46. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018;67:901–4. https://doi.org/10.1093/sysbio/syy032.
47. Drummond AJ, Rambaut A. BEAST: bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214. https://doi.org/10.1186/1471-2148-7-214.
48. Lemey P, Minin VN, Bielejec F, Kosakovsky Pond SL, Suchard MA. A counting renaissance: combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection. Bioinformatics. 2012;28:3248–56. https://doi.org/10.1093/bioinformatics/bts580.
49. RStudio Team. RStudio: Integrated Development for R. RStudio. Boston, MA: PBC; 2020.
50. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol. 2020;6:veaa061. https://doi.org/10.1093/ve/veaa061.
51. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2:vew007–007. https://doi.org/10.1093/ve/vew007.
52. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol Biol Evol. 2006;23:7–9. https://doi.org/10.1093/molbev/msj021.
53. Ghafari M, du Plessis L, Raghwani J, Bhatt S, Xu B, Pybus OG, et al. Purifying selection determines the short-term time dependency of Evolutionary Rates in SARS-CoV-2 and pH1N1 influenza. Mol Biol Evol. 2022;39:msac009. https://doi.org/10.1093/molbev/msac009.
54. Stefan AM, Gronau QF, Schönbrodt FD, Wagenmakers E-J. A tutorial on Bayes Factor Design Analysis using an informed prior. Behav Res Methods. 2019;51:1042–58. https://doi.org/10.3758/s13428-018-01189-8.
55. Rahman FN, Rahman AKMF, Iwuagwu AO, Dalal K. COVID-19 transmission due to Mass Mobility before and after the Largest Festival in Bangladesh: an epidemiologic study. Inquiry. 2021;58:469580211023464. https://doi.org/10.1177/00469580211023464.
56. The New York Times. The Trouble With the Case Curve During the Holidays n.d. https://www.nytimes.com/interactive/2021/11/22/us/covid-data-holiday-averages.html
57. Brum E, Saha S, Sania A, Tanmoy AM, Hooda Y, Tanni AA et al. Surging COVID-19 in Bangladesh driven by B.1.351 variant. MedRxiv 2021:2021.04.27.21255706. https://doi.org/10.1101/2021.04.27.21255706
58. Gu H, Xie R, Adam DC, Tsui JL-H, Chu DK, Chang LDJ, et al. Genomic epidemiology of SARS-CoV-2 under an elimination strategy in Hong Kong. Nat Commun. 2022;13:736. https://doi.org/10.1038/s41467-022-28420-7.
59. Geoghegan JL, Ren X, Storey M, Hadfield J, Jelley L, Jefferies S, et al. Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. Nat Commun. 2020;11:6351. https://doi.org/10.1038/s41467-020-20235-8.
60. Islam MT, Talukder AK, Siddiqui MN, Islam T. Tackling the COVID-19 pandemic: the Bangladesh perspective. J Public Health Res. 2020;9:1794. https://doi.org/10.4081/jphr.2020.1794.
61. Di Giallonardo F, Duchene S, Puglia I, Curini V, Profeta F, Cammà C, et al. Genomic epidemiology of the First Wave of SARS-CoV-2 in Italy. Viruses. 2020;12. https://doi.org/10.3390/v12121438.
62. Islam J, Guo X, Ali MA, Islam MA, Qi X, Zhuang G. Spatial pattern of COVID-19 in Bangladesh: an ecological study. BMJ Open. 2021;11:e047566. https://doi.org/10.1136/bmjopen-2020-047566.
63. Rabbi MFA, Khan MI, Hasan S, Chalita M, Hasan KN, Sufian A et al. Large scale genomic and evolutionary study reveals SARS-CoV-2 virus isolates from Bangladesh strongly correlate with European origin and not with China. BioRxiv 2021:2021.01.17.425424. https://doi.org/10.1101/2021.01.17.425424
64. Rashid SR. Remittance practices in rural Bangladesh: a gendered analysis. Migr Dev. 2022;11:21–39. https://doi.org/10.1080/21632324.2019.1697494.
65. IOM, Bangladesh. POINT OF ENTRY ANALYSIS COVID-19. 2021.
66. Sayeed MA, Ferdous J, Saha O, Islam S, Choudhury SD, Abedin J, et al. Transmission Dynamics and genomic epidemiology of emerging variants of

SARS-CoV-2 in Bangladesh. Trop Med Infect Dis. 2022;7. https://doi.org/10.3390/tropicalmed7080197.
67. Hasan MS, Islam MT, Alam ASMRU, Sarkar SL, Rahman MS, Islam OK, et al. Initial reports of the SARS-CoV-2 Delta variant (B.1.617.2 lineage) in Bangladeshi patients: risks of cross-border transmission from India. Health Sci Rep. 2021;4:e366. https://doi.org/10.1002/hsr2.366.
68. Sawadogo Y, Galal L, Belarbi E, Zongo A, Schubert G, Leendertz F, et al. Genomic epidemiology of SARS-CoV-2 in Western Burkina Faso, West Africa. Viruses. 2022;14. https://doi.org/10.3390/v14122788.
69. Butera Y, Mukantwari E, Artesi M, Umuringa J, d'arc, O'Toole ÁN, Hill V, et al. Genomic sequencing of SARS-CoV-2 in Rwanda reveals the importance of incoming travelers on lineage diversity. Nat Commun. 2021;12:5705. https://doi.org/10.1038/s41467-021-25985-7.
70. Lee JN, Mahmud M, Morduch J, Ravindran S, Shonchoy AS. Migration, externalities, and the diffusion of COVID-19 in South Asia. J Public Econ. 2021;193:104312. https://doi.org/10.1016/j.jpubeco.2020.104312.
71. Rahaman KR, Mahmud MS, Mallick B. Challenges of Testing COVID-19 cases in Bangladesh. Int J Environ Res Public Health. 2020;17. https://doi.org/10.3390/ijerph17186439.
72. Brown TS, Engø-Monsen K, Kiang MV, Mahmud AS, Maude RJ, Buckee CO. The impact of mobility network properties on predicted epidemic dynamics in Dhaka and Bangkok. Epidemics. 2021;35:100441. https://doi.org/10.1016/j.epidem.2021.100441.
73. Financial Express. Commuting to Dhaka -- the options n.d. https://thefinancialexpress.com.bd/views/opinions/commuting-to-dhaka-the-options-1581178251
74. González-Leonardo M, Rowe F, Fresolone-Caparrós A. Rural revival? The rise in internal migration to rural areas during the COVID-19 pandemic. Who moved and where? J Rural Stud. 2022;96:332–42. https://doi.org/10.1016/j.jrurstud.2022.11.006.
75. Miller D, Martin MA, Harel N, Tirosh O, Kustin T, Meir M, et al. Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. Nat Commun. 2020;11:5518. https://doi.org/10.1038/s41467-020-19248-0.
76. Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Sci (1979). 2020;369:1255–60. https://doi.org/10.1126/science.abd2161.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.