

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

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

8
9 ¹Department of Pathobiology and Population Sciences, Royal Veterinary College (RVC), Hatfield,
10 Hertfordshire, United Kingdom

11 ²Institute of Ecology and Evolution, University of Edinburgh, King's Buildings, Edinburgh, United
12 Kingdom

13 ³Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle,
14 Washington, United States of America

15 ⁴Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

16 ⁵Swiss Institute of Bioinformatics, Lausanne, Switzerland

17 ⁶Bangladesh Livestock Research Institute (BLRI), Ashulia, Dhaka, Bangladesh

18 ⁷Global Change Center, Virginia Tech, Blacksburg, Virginia, United States of America

19 ⁸Chattogram Veterinary and Animal Sciences University (CVASU), Khulshi, Chattogram, Bangladesh

20 ⁹School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University, Wagga Wagga,
21 New South Wales, Australia

22 ¹⁰NSU Genome Research Institute (NGRI), North South University, Bashundhara, Dhaka, Bangladesh

23 ¹¹Department of Biochemistry and Microbiology, North South University, Bashundhara, Dhaka,
24 Bangladesh

25 ¹²Bangladesh Reference Institute for Chemical Measurements (BRiCM), Dhanmondi, Dhaka, Bangladesh

26 ¹³National Institute of Laboratory Medicine and Referral Centre (NILMRC), Agargoan, Dhaka,
27 Bangladesh

28 ¹⁴Centers for Disease Control and Prevention (CDC), Atlanta, Georgia, United States of America

29 ¹⁵United Kingdom Health Security Agency (UKHSA), Porton Down, Salisbury, United Kingdom

30 ¹⁶Université de Lyon, INRAE, VetAgro Sup, UMR EPIA, Marcy l'Etoile, France

31 ¹⁷Université Clermont Auvergne, INRAE, VetAgro Sup, UMR EPIA, Saint Genes Champanelle, France

32 ¹⁸National Institute of Preventive and Social Medicine (NIPSOM), Ministry of Health and Family Welfare,
33 Dhaka, Bangladesh

34 *Authors contributed equally to this work.

35 Correspondence to: lcarnegie20@rvc.ac.uk; sahill@rvc.ac.uk; msamad@blri.gov.bd.

36

37 **Abstract**

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

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

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

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

51 **Key words:** SARS-CoV-2, Phylodynamics, Bangladesh, Genomic epidemiology

52

53 **1. Background**

54 Viral genomic analyses of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have provided
55 insight into viral spread on both local and global scales and helped to inform infection control measures
56 [1–6]. For instance, genomic analysis has been used to identify that Brazil shifted from being a net
57 importer to an exporter of SARS-CoV-2 lineages over the course of 2020 [5], and to reconstruct patterns

58 of within-country viral movements in France during the country's first two Coronavirus disease 2019
59 (COVID-19) pandemic waves [7]. Using viral genomic data allows for epidemiological insight that cannot
60 be captured from case data alone, including the inference of transmission dynamics prior to the first
61 recorded sample or during periods with limited disease surveillance.

62

63 Despite often having high age-specific infection fatality rates (IFRs) [8], limited genomic studies of SARS-
64 CoV-2 transmission dynamics have been conducted in the world's least developed countries (LDCs) [8-
65 10]. LDCs are low/middle-income countries (LMICs) that the United Nations (UN) classifies as possessing
66 particularly high levels of poverty, human resource weakness, and economic vulnerability [10,11].
67 Instead, most studies focus on findings from high income countries (HICs), particularly in Western
68 Europe and North America, likely reflecting a greater ability to invest in both sequence generation and
69 research efforts [8,9,12]. This is problematic as findings from HICs may not be readily applicable to
70 LDCs. LDCs often face unique challenges in controlling the virus, including limited public health
71 resources, weak healthcare systems, and large sections of their populations highly susceptible to
72 external economic pressures [8,9,13]. The indiscriminate application of control measures that are
73 effective in high-income countries (HICs) may lead to unintended consequences in LDCs, particularly in
74 settings where enforcement is challenging [13]. For example, stay-at-home orders have at times led to
75 mass economic migration from cities, resulting in virus exports to larger geographic areas [14].
76 Identifying factors that influence transmission dynamics in LDCs allows for better tailoring of control
77 measures to local epidemiological contexts, thereby increasing their chances of success [13,15].

78

79 The limited number of genomic epidemiological analyses in LDCs is exemplified in Bangladesh, where,
80 despite some insights, many aspects of SARS-CoV-2 spread remain poorly understood. Cowley et al.,
81 (2021) compared observed mobility pattern data with time-scaled viral phylogenies to suggest that
82 mass migration out of Dhaka, the capital of Bangladesh, may have driven early viral spread across the
83 country [14]. However, in contrast to many HICs, the potential association between spatial and

84 demographic factors and viral dispersal patterns has not been quantified [14]. Likewise, while studies
85 reported that SARS-CoV-2 may have appeared in Bangladesh weeks before the first case report [14]
86 and was likely introduced multiple times from several countries [16–18], the number of viral lineage
87 imports to and exports from Bangladesh have not been estimated.

88

89 Bangladesh is the most densely populated country in the world that is not a city-state (165 million
90 inhabitants living at over 1000 people/km²), and one of several LDCs in South Asia [14,19–23]. The first
91 case of SARS-CoV-2 in Bangladesh was confirmed on March 8th, 2020 [19,24]. In response to rising cases,
92 a nationwide lockdown (known as a ‘National General Holiday’) was introduced on the 26th of March
93 and lasted until the 30th of May. Restrictions and screening procedures were implemented at
94 educational institutions, workplaces and international airports [23,25]. Infections nevertheless
95 increased steeply in April, with nearly 313,000 confirmed cases reported by 31st August 2020 (a total of
96 4,281 deaths) [18]. Following the end of the national lockdown, a subsequent zone-coded lockdown
97 was used whereby movement within and between certain high-incidence areas was restricted [23,25–
98 27].

99

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

107

108 **2. Methods**

109

110 2.1. Sample Collection

111

112 Samples were chosen for sequencing from residual anonymised diagnostic samples collected between
113 26th April - 31st August 2020 that were received by the National Reference Laboratory for Avian
114 Influenza, Bangladesh Livestock Research Institute (BLRI) and that had tested positive by Real-time
115 reverse transcription polymerase chain reaction (rRT-PCR). There was no patient or public involvement
116 in the design of this study. Samples were nasopharyngeal or oropharyngeal swabs or both and were
117 collected along the clinical criteria according to the National Guidelines on Clinical Management of
118 Coronavirus Disease 2019 (COVID-19) published by the Directorate General of Health Services (DGHS),
119 Bangladesh. A structured data collection questionnaire (**Figure S1**) was collected alongside all samples,
120 detailing each patient's geographic origin, clinical symptoms, travel history, and probable viral source.
121 rRT-PCR for SARS-CoV-2 detection was performed using fluorescent probes and the result was
122 considered positive when the cycle threshold (Ct) values of both the ORF1ab and N genes were <32 and
123 <35, as measured by ROX [28] and FAM [29] dyes [30–32]. We attempted to choose samples
124 proportionally to confirmed cases per district (see **Table S1** for a comparison of sequences to confirmed
125 cases per district), whilst also still including a high percentage of samples from likely under-sampled
126 regions. We preferentially selected samples for which information on travel history (including
127 confirmation of absence of travel) was present.

128

129 2.2 Sequencing

130

131 We sequenced 175 whole SARS-CoV-2 genomes from samples using a Nanopore GridION and the nCoV-
132 2019 sequencing protocol v3 (LoCost) V.3 incorporating the ARTIC multiplex-PCR sequencing pipeline
133 (as described in [33]). We combined newly generated sequences with 140 Bangladeshi sequences
134 accessed (July 2022) via the GISAID [34] EpiCOV database (www.gisaid.org) collected from patients

135 during the same sampling period. These included: (i) sequences already used in published papers (n =
136 139) and (ii) unpublished sequences where submitting laboratories granted us permission for their
137 inclusion (n = 1). All SARS-CoV-2 sequences were aligned using Llama (Local Lineage and Monophyly
138 Assessment) [35] v0.1 with default parameters. Using AliView [36] v1.28 we confirmed that there were
139 no sequences from the alignment that either were duplicated, short (<70% of the total sequence
140 length), or indicative of containing sequencing or assembly errors.

141

142 **2.3. Epidemiological Data**

143

144 We obtained COVID-19 confirmed case data between 26th April - 31st August 2020 from the Institute of
145 Epidemiology Disease Control and Research (IEDCR) “COVID-19 Dynamic Dashboard for Bangladesh”
146 [37]. Aggregated daily case counts were available for each of Bangladesh’s eight divisions (Chattogram,
147 Dhaka, Sylhet, Khulna, Barishal, Rajshahi, Rangpur, Mymensingh). For the national-level data, days in
148 which data was missing from one or more divisions were excluded (n = 10) from analyses.

149

150 **2.4. Epidemiological Analyses**

151

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

158

159 **2.5. Transmission Lineage Analyses**

160

161 We first assembled a global alignment by sampling SARS-CoV-2 genome sequences evenly for each
162 week and country (oldest sequence: 2019-12-24, newest sequence: 2020-08-31). Temporal outliers,
163 defined as samples that fell beyond 4 interquartile ranges of the expected divergence given were
164 identified and removed using TreeTime [40] and a fixed clock rate of 0.00075 substitutions/site/year.
165 This resulted in a dataset of 5804 sequences. We then used ThorneyBEAST
166 (https://beast.community/thorney_beast) to estimate a posterior molecular clock tree using a pipeline
167 similar to that of du Plessis et al., (2021) [41], Raghwani et al., (2022) [4] and Gutierrez et al., (2021)
168 [42]. This employed a strict molecular clock model [43] (0.00075 substitutions/site/year), and a Skygrid
169 coalescent tree prior [44]. For the starting tree, we used a maximum-likelihood tree produced with IQ-
170 TREE [45] under a Jukes-Cantor substitution model, using Wuhan/WH04/2020 as an outgroup. During
171 the Markov chain Monte Carlo (MCMC), we constrained our tree search to sample only node height
172 and resolutions of polytomies present in the starting tree. We completed 5 chains of 200 million steps
173 with sampling every 1.8 million steps, removing the first 10% of steps as burn-in. The resulting sample
174 of 500 trees from the posterior were used to generate a maximum clade credibility (MCC) tree. We
175 confirmed that the Effective Sample Size (ESS) of all parameters was >200 using Tracer [46] v1.7.1.

176
177 We then conducted a discrete trait analysis (DTA) with these 500 empirical trees in BEAST [47] to
178 estimate ancestral node locations (Bangladesh or non-Bangladesh). We used a robust counting
179 approach [48] to evaluate the expected number of location state transitions into and out of Bangladesh
180 across a posterior sample of 2000 trees.

181
182 Finally, we used the location-annotated MCC tree to identify transmission lineages in Bangladesh,
183 following previously described methods [41]. Briefly, nodes with a posterior probability > 0.5 inferred
184 to be located in Bangladesh are considered “Bangladeshi nodes”. A depth-first search initiated at a
185 random Bangladeshi node is run until complete or until a non-Bangladeshi node is encountered, with
186 encountered nodes labelled as belonging to the same transmission lineage. A new starting node is

187 chosen until every Bangladeshi node is visited, and hence all Bangladeshi tips have been assigned a
188 lineage or classified as a singleton. For each Bangladeshi lineage, we estimated time to the most recent
189 common ancestor (TMRCA).

190

191 2.6. Discrete Trait Analysis with Generalized Linear Model

192

193 To investigate how viral dispersal was influenced by spatial and demographic factors in Bangladesh we
194 focused on the two largest transmission lineages identified in the global dataset (**Material and**
195 **Methods Section 2.5.**) Only those whole-genome sequences with known sampling districts were used
196 ($n = 142/307$ for lineage 2 and $n = 52/97$ for lineage 8, remaining total with sampling districts = 194)
197 (**Tables S2-3**). To facilitate effective discrete phylogeography analyses, we reduced the original 28
198 districts (second administrative level), covering six out of the eight Bangladeshi divisions (first
199 administrative level) into a smaller set of discrete locations. In brief, we employed k-means clustering
200 based on Euclidean distance, as implemented in the '*stats*' package in R [49] v4.1.2, to group these
201 sequences into eight geographic regions, each representing a grouping of several districts. All
202 sequences were assigned to their respective district centroid coordinates for this purpose. We
203 assembled data on chosen predictors for each of these eight geographical groups of districts:
204 population size and density, the ratio between the number of cases and population size, and geographic
205 distance (ellipsoid distance between grouping centroids) (see **Tables S4-5** for details). Whilst we did
206 not expect a strong temporal signal given the low mutation rate of SARS-CoV-2 and the short timescale
207 of our study (e.g., [14,50]), we used TempEst [51] v1.5.3 to check that the temporal signal in each
208 lineage alignment was not incompatible with using fixed substitution rates based on the literature and
209 to check for outlier sequences (**Figure S2**). We used a DTA model implemented in BEAST [47] v1.10.4
210 to reconstruct virus lineage movements between the eight pre-defined geographic groupings. A strict
211 molecular clock model with a fixed rate (0.001 substitutions/site/year) and an SRD06 substitutional
212 model [52] was shared between both transmission lineages to improve convergence, with a separate

213 exponential growth coalescent tree prior that allows for different timings of lineage growth. We used
214 a faster clock rate (0.001 substitutions/site/year) in these shorter timescale phylogeographic analyses,
215 compared to the previous global transmission lineages analyses (0.00075 substitutions/site/year), to
216 account for the existence of slightly deleterious mutations circulating in the Bangladesh transmission
217 lineages that had not been removed by purifying selection [53]. We used separate generalised-linear
218 models (GLMs) to determine which covariate(s) (**Table S4-5**) best predicted the frequency of viral
219 lineage movements between locations for each lineage. The migration rates in the DTA model are
220 parameterised as a GLM using covariates as coefficients of the GLM. DTA-GLM analyses that included
221 either population density or population size predictors were run separately because of the multi-
222 collinearity between these predictors. For each analysis, we executed two chains of 150 million steps
223 logged every 15,000 steps, removing the first 10% of steps as burn-in. We used Tracer [46] v1.7.1 to
224 confirm that the ESS was >200 for all parameters. The support of the Bayes Factor (BF) for transmission
225 between discrete traits was interpreted as described previously in [54], with BF > 100 ‘very strong
226 support’, and BF > 10 indicating ‘strong support’.

227

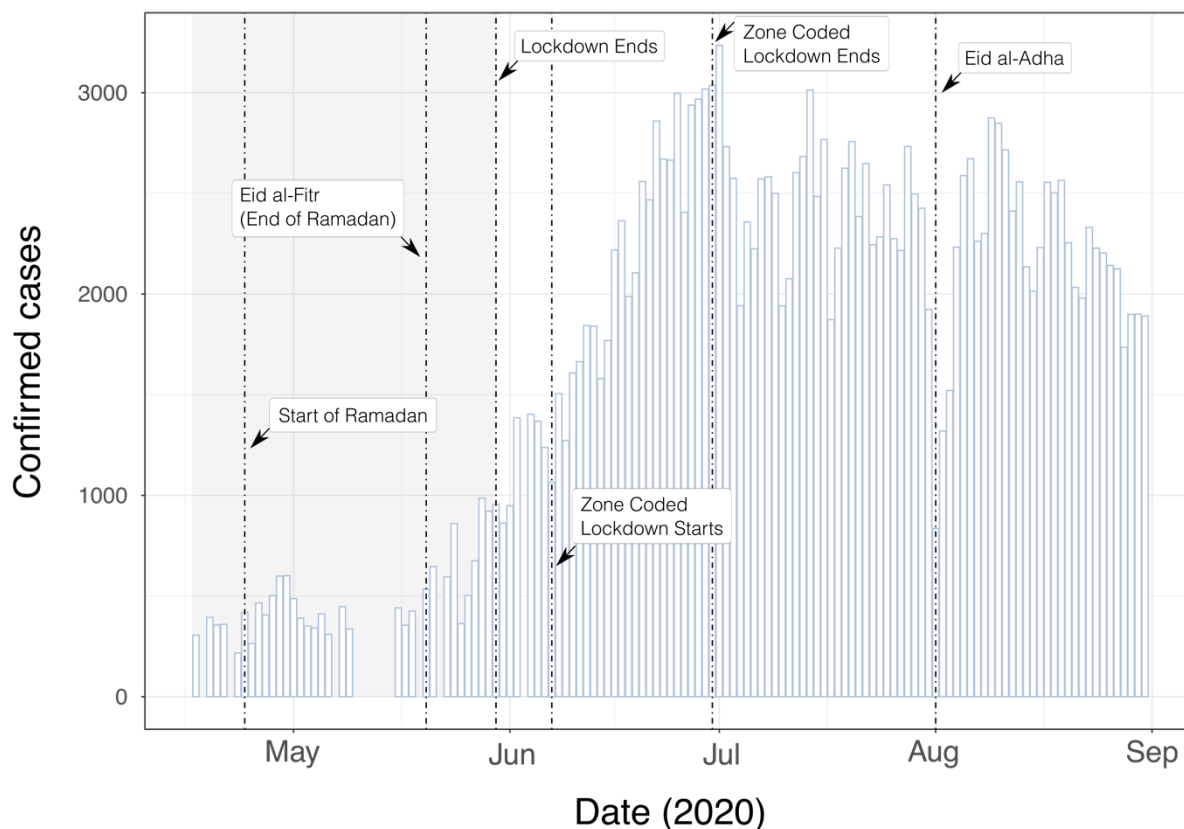
228 **3. Results**

229

230 **3.1. Epidemiological Findings**

231

232 COVID-19 confirmed cases increased towards the end of the national lockdown and continued rising
233 during the subsequent zone-coded lockdown, peaking at the start of July 2020 (**Figure 1**). Reported
234 case numbers dropped sharply during Eid al-Adha, which is one of the most important Muslim festivals
235 and marked by a period of public holiday in Bangladesh [55] (31st July 2020 – 1st August 2020) (**Figure**
236 **1**). This is consistent with reduced reporting and testing in other countries in periods surrounding
237 important religious events [56].



238

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

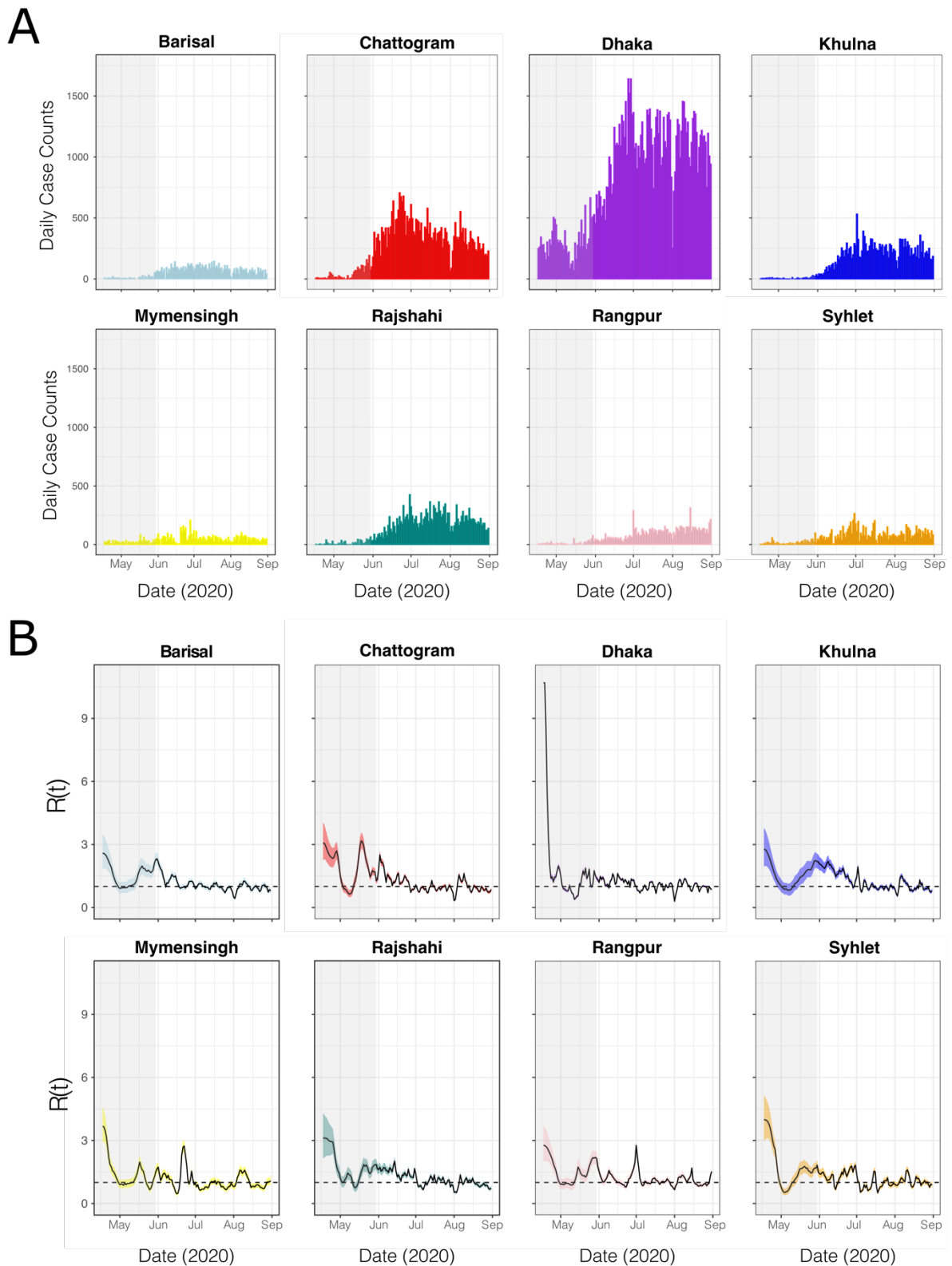
241

242 We plotted daily COVID-19 confirmed case counts in Bangladesh’s eight divisions between 17th April
 243 2020 and 31st August 2020. Case counts in each division were similar to the pattern observed at the
 244 country level (**Figures 1 & 2A**). It is possible that rising cases could be somewhat driven by increased
 245 testing capacity over time [14]. In particular, during the first months of the pandemic, residents in
 246 Dhaka, the country’s largest and most populated city, had easier access to SARS-CoV-2 testing
 247 compared to other regions in Bangladesh [14,57]. Consequently, the Dhaka division reported the
 248 majority (50.5%) of confirmed SARS-CoV-2 cases in the country (**Figure 2A**) [14,57]. Spikes in reported
 249 cases during the first wave of the pandemic were observed in some other divisions, e.g., in Rangpur in
 250 the start of July 2020 (**Figure 2A**).

251

252 Division-specific effective reproduction numbers $R(t)$ showed similar temporal trends in most
253 Bangladeshi divisions (**Figure 2B**). $R(t)$ typically fell in the first half of the national lockdown, but then
254 peaked above 1 near the middle or end of May, concurrent with both the religious festival of Eid al-Fitr
255 and the end of the national lockdown. $R(t)$ then slowly decreased and fluctuated around 1 by the start
256 of July (**Figure 2B**). Division-specific spikes and troughs in $R(t)$ were also observed, such as a reduction
257 of $R(t)$ below 1 in mid-June in Mymensingh (**Figure 2B**).

258



259

260 *Figure 2. Grey-highlighted regions indicate a national lockdown. A) Daily COVID-19 confirmed cases in*

261 *each division. B) Effective reproduction number ($R(t)$) by division. Median R_t estimates are shown by*

262 *black lines, and 95% Bayesian credible intervals by coloured ribbons.*

263 3.2. Transmission Lineage Findings

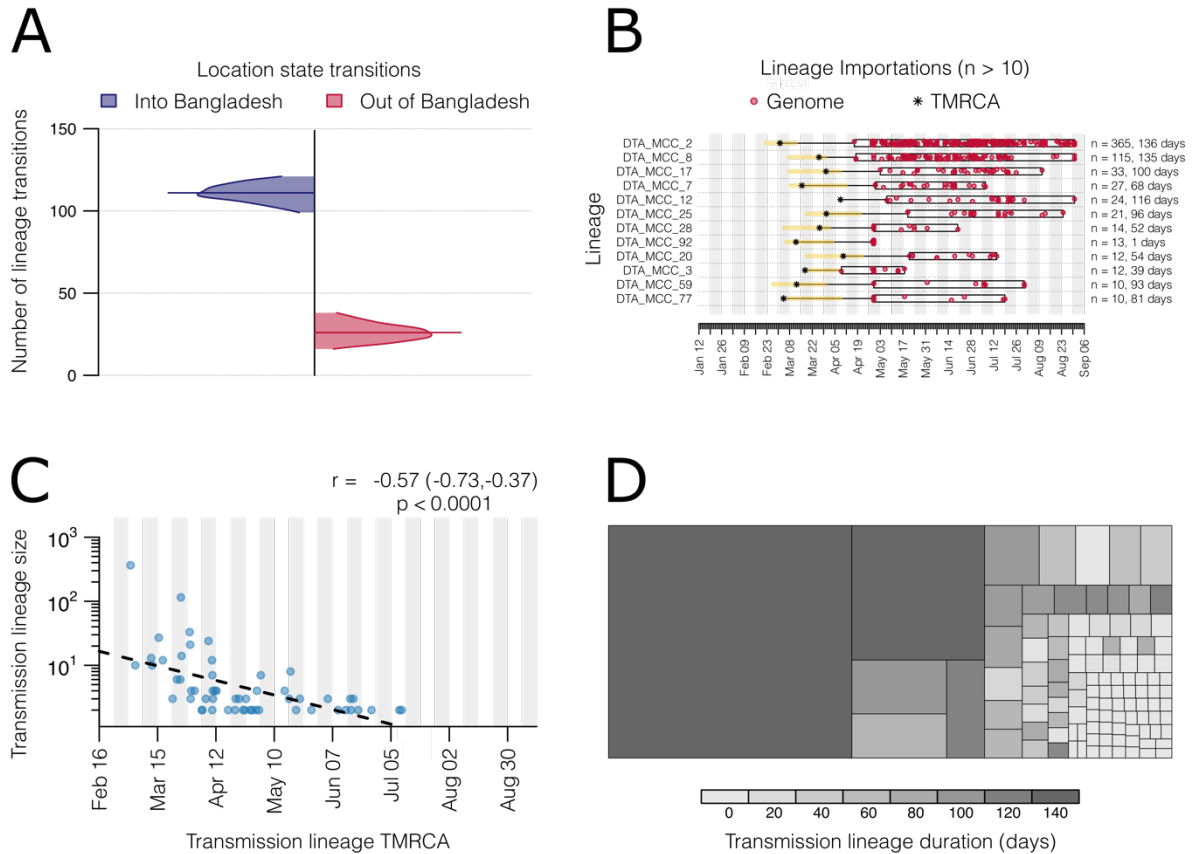
264

265 We estimated the number of SARS-CoV-2 lineage importations over time to Bangladesh. The estimated
266 number of viral lineage movements into the country ($n = 111$, 95% HPD: 99 – 121) was approximately
267 4 times higher than the number of viral lineage movements from Bangladesh to other countries globally
268 ($n = 26$, 95% HPD: 16 – 38) (**Figure 3A**), indicating that Bangladesh was mainly an importer of virus
269 lineages during the first wave.

270

271 We detected 56 distinct Bangladesh transmission lineages (i.e., lineages spreading within Bangladesh,
272 represented by >1 genome), and 51 singletons (singular Bangladesh genomes representing lineage
273 imports originating from non-Bangladesh regions, yet without evidence of autochthonous
274 transmission) (**Figure S3**). Twelve lineages contained >10 sequences, and two lineages contained >100
275 genomes. Several lineages ($n = 16$, 28.6%), including the two largest lineages, had an estimated TMRCA
276 that fell either prior to or within the following week of the start of the national lockdown on the 26th
277 March (**Figures 3B & S3**) [25]. Multiple lineages, however, had TMRCA's overlapping with the period of
278 nationwide lockdown (**Figures 3C & S3**). Finally, there was a strong negative exponential relationship
279 between the estimated TMRCA and transmission lineage size (**Figure 3C**). Larger transmission lineages
280 in Bangladesh were generally associated with longer sampling durations (**Figure 3D**).

281



282

283 **Figure 3. SARS-CoV-2 transmission lineage characteristics in Bangladesh. A)** Number of location state

284 transitions between the phylogenetic traits Bangladesh/Global (imports into Bangladesh = blue, exports

285 from Bangladesh = red), as detected via the robust counting approach. Posterior distributions are

286 truncated at their 95% highest posterior distribution (HPD) interval limits and median estimates are

287 shown using horizontal lines. **B)** Duration and timing of the largest Bangladesh transmission lineages (>

288 10 genomes). Each row represents a transmission lineage, and red dots indicate genome sampling times.

289 Boxes and labels on the right axis show the sampling duration (see **Figure S3** for more details on

290 sampling duration per lineage), and number of sampled genomes (n). Asterisks show the median

291 estimated time to most recent common ancestor (TMRCA) for each lineage, with the 95% HPD as a

292 yellow bar **C)** Relationship between transmission lineage size and TMRCA, with a dashed line indicating

293 the slope of a linear regression. The Pearson correlation coefficient, 95% confidence interval, and p-value

294 are shown. **D)** Cells represent transmission lineages and singletons, each coloured by estimated duration

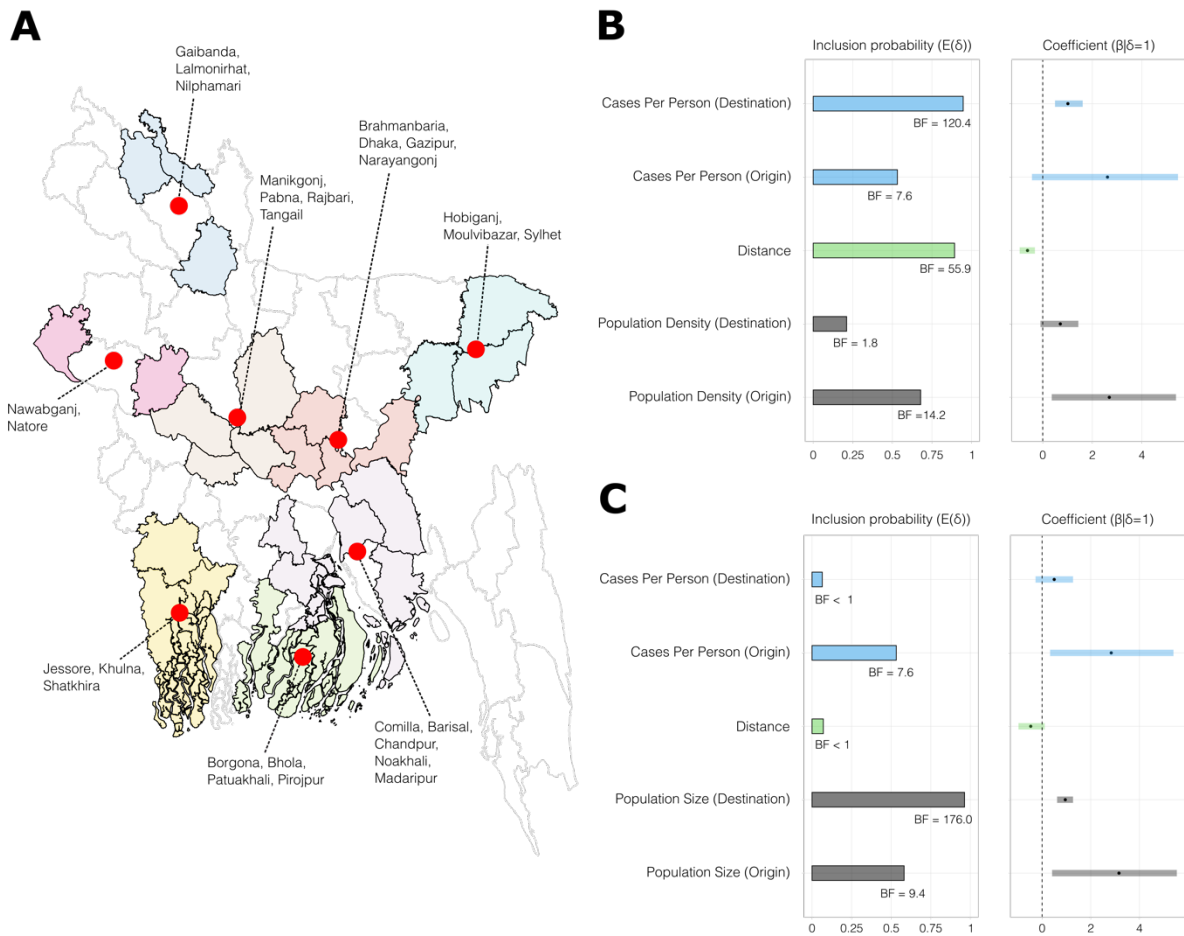
295 and scaled by size.

296

297 3.3. Drivers of virus spread within Bangladesh

298

299 We used DTA models coupled with GLMs to identify significant predictors of virus lineage dissemination
300 among eight geographical groups each containing districts from multiple Bangladeshi divisions (**Figure**
301 **4A**), using only those sequences from the two largest Bangladesh transmission lineages (**Figure 3**). We
302 fitted two models with either population size or population density as the demographic predictor
303 (covariate values shown in **Table S5 & Figure S4**). We observed consistent directionality and effect sizes
304 for two population metrics in each of their respective models, but notable differences were found in
305 the BF support and inclusion probabilities. In the model including population density as a predictor
306 (**Figure 4B**), there was strong evidence ($BF > 10$, following: [54]) that virus lineages were more likely to
307 move from areas with high population density. We also found very strong support ($BF > 100$) for virus
308 movement between geographically close locations and very strong support ($BF > 100$) towards areas
309 with a high number of confirmed cases per capita. In the model including population size as predictor
310 (**Figure 4C**), there was strong support for virus spread occurring more commonly to and from areas
311 with high population sizes, and for virus movement towards locations with a high number of confirmed
312 cases per capita. However, distance was not identified as a significant predictor ($BF < 1$).



313

314 **Figure 4. Factors associated with SARS-CoV-2 spread. A)** Map of Bangladesh showing the eight

315 geographical groups of districts used as discrete traits in the DTA-GLM. Region centroids are marked by

316 red dots. **B and C)** Predictors of SARS-CoV-2 spread based on models with either population density (**B**)

317 or population size (**C**) included as predictors. Bar and line colours indicate different covariates, with

318 origin and destination predictor of a covariate given the same colour within each plot. Inclusion

319 probability is the posterior expectation that the indicator variable is associated with each predictor $E(\delta)$

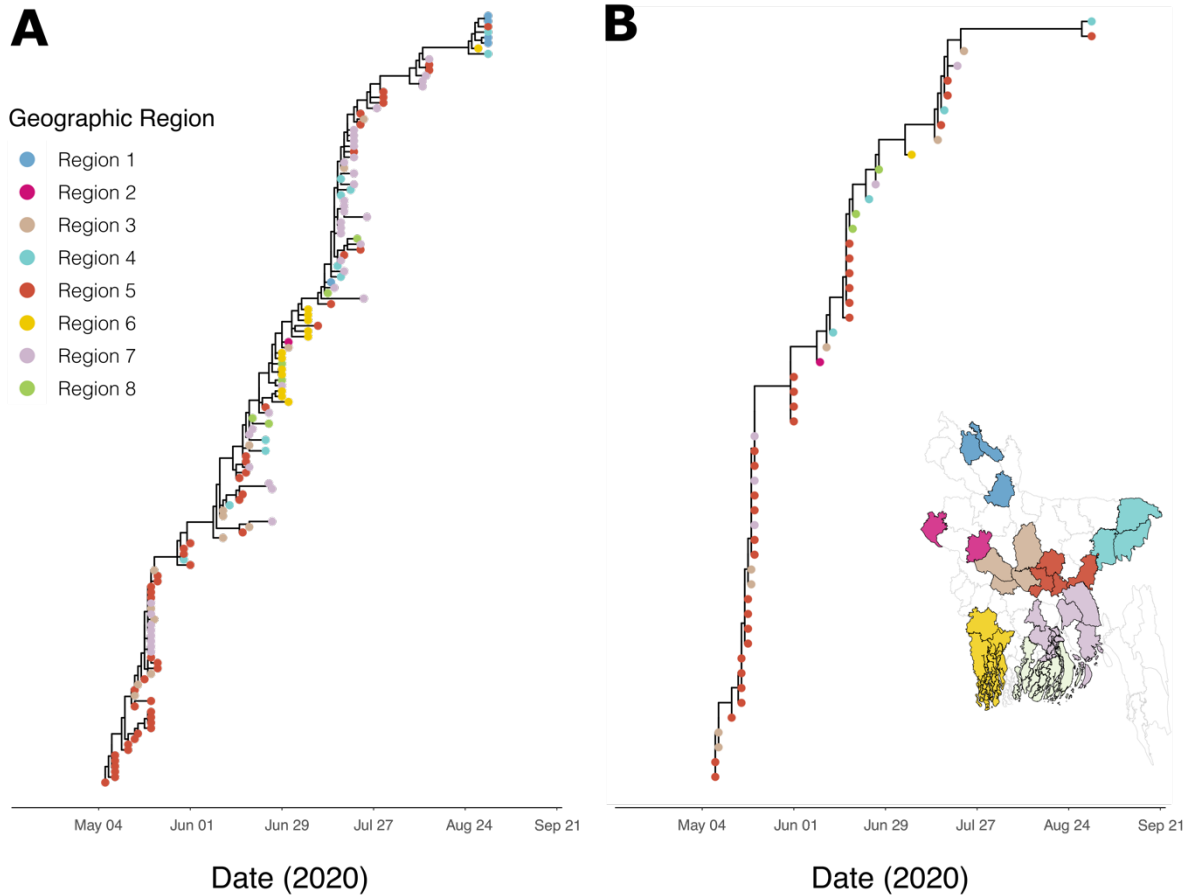
320 and suggests that the predictor is associated with different rates of viral diffusion. Bayes Factor (BF)

321 support values for each covariate are indicated by black text annotations. The coefficient ($\beta|\delta=1$)

322 represents the contribution of each predictor on a log scale when the predictor is included in the model,

323 with the 95% credible interval of the GLM coefficients (β) represented by horizontal lines.

324



325

326 **Figure 5.** MCC phylogenies of the two largest lineages detected. **A)** Lineage 2, and **B)** Lineage 8. Branch
 327 lengths represent time, as shown on the axis. Tips are coloured by the sampling geographic region used
 328 in the DTA-GLM analyses, as shown in the inset map.

329

330 4. Discussion

331

332 Genomic epidemiology has played a crucial role in understanding the global and regional movement of
 333 SARS-CoV-2. However, research efforts have disproportionately neglected studying virus spread within
 334 the world's least developed countries. To contribute to addressing this disparity, we use a combination
 335 of epidemiological and phylodynamic analyses to investigate the transmission dynamics of SARS-CoV-2
 336 in Bangladesh during the country's initial wave of COVID-19. We determined that $R(t)$ showed a
 337 declining trend during the start of the national lockdown but rose slightly towards the end of the

338 national lockdown in late May 2020, which was perhaps driven by activity surrounding Eid al-Fitr.
339 Subsequently, $R(t)$ slowly stabilised and fluctuated around ~ 1 in all divisions. Our study revealed that
340 cases in Bangladesh were initiated by more than 50 instances of virus introduction into the country,
341 with multiple lineages maintained even during the period of national lockdown. Finally, we identified
342 using discrete phylogeographic analyses that patterns of virus dispersal may have been shaped by
343 population density and/or size, as well as geographical proximity, although the latter was not
344 consistently important across all analyses.

345

346 The detection of over 50 virus introductions during Bangladesh's initial COVID-19 wave, and of multiple
347 persistent transmission lineages during the lockdown, contrasts with studies from Hong Kong [58] and
348 New Zealand [59], where virus cases were linked to repeated individual introductions rather than
349 persistently present lineages. Both our study and the Hong Kong study were based on similar pipelines
350 adapted from du Plessis et al. (2021) [41], and therefore the observed difference is likely not caused by
351 different study designs. Instead, the distinction likely reflects a greater inability to adequately maintain
352 social distancing in Bangladesh's high-density urban areas and the relatively greater capacity of these
353 other countries to control local transmission through measures such as community surveillance and
354 contact tracing [12,44]. Additionally, New Zealand and Hong Kong adopted stringent inbound travel
355 regulations and testing measures [58,59]. In contrast, Bangladesh had limited thermal scanners at
356 Dhaka international airports, and reported low quarantine compliance [19,60]. These differences may
357 have played a further role in the observed variations in SARS-CoV-2 lineage dynamics between
358 locations.

359

360 In congruence with previous studies in South Africa [6] and Italy [61], we found that the estimated
361 number of virus importations into the country considerably outweighed exports. The return of a large
362 number of Bangladeshi nationals to the country during the early stages of the pandemic may have
363 allowed for extensive lineage importation to the country from a wide range of geographical locations

364 [14,16–18,62]. Notably, over 10 million Bangladeshi citizens live abroad, and the number of migrant
365 labourers returning to the country was 8 times higher in 2020 than in previous years [18,63,64].
366 Introductions may have been seeded via international airports [14,16–18,60], as well as through cross-
367 border human mobility with neighbouring countries (e.g., India [65–67]). The latter has been an
368 important pathway for the introduction of SARS-CoV-2 lineages to other LDCs (e.g., Burkina Faso [68]
369 and Rwanda [69]). However, as the number and origin of introductions and exports estimated from
370 genomic data are dependent on sample size [41], our analyses may be biased towards capturing
371 imports rather than exports, given relatively less intense sampling in Bangladesh compared to many
372 other regions globally.

373

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

388

389 In one of our discrete phylogeographic analyses, we identified that viral movement occurred
390 preferentially between more proximate geographic groupings of districts in Bangladesh. This suggests
391 that certain lockdown measures (e.g., travel advisories, railway closures, domestic flight suspensions
392 [23,25–27]), likely had a more substantial impact on restricting long-distance human mobility and virus
393 spread, compared to movement and viral dispersal between closer or neighbouring divisions. Distance
394 has also been identified as a predictor of SARS-CoV-2 spread in several other countries, including Israel
395 and Brazil [75,76]. That said, the presence of strong evidence (using BF support) for distance depended
396 here on whether human population density or size was considered, with only strong evidence detected
397 when the population density predictor was also included in the model. This could indicate that distance
398 had a relatively weak effect on viral transmission dispersal between regions in the country, perhaps
399 because Bangladesh is relatively small and its population is highly mobile [72].

400

401 Our study has several limitations. Firstly, only a very small fraction of the total confirmed SARS-CoV-2
402 cases could be sequenced due to resource constraints. Greater sampling and sequencing of cases would
403 have allowed us to analyse the virus dispersal patterns in Bangladesh at higher resolution. Second, the
404 limited genomic surveillance in certain locations outside of Dhaka where many cases were confirmed,
405 particularly the absence of sequences from Chattogram, the country's second-largest city, indicates
406 that our analyses may have underestimated the extent and frequency of virus movements within
407 Bangladesh. As such, the outputs of our DTA-GLM models may have not fully captured the strength or
408 impact that certain demographic predictors had on within-country viral dispersal [14]. We attempted
409 to mitigate this limitation by choosing samples for sequencing roughly proportionally to the confirmed
410 cases per Bangladeshi district while retaining most samples from any particularly under sampled
411 locations. However, for several regions, very few samples were available and case reporting is likely to
412 be inherently spatially biased. Third, due to the short time-period of our study, we were unable to
413 investigate changes in the drivers of viral spread during the study period. The dynamics of transmission
414 and the sources of infection may have altered over time, especially in response to mobility restrictions

415 and interventions such as border closures. Furthermore, the geographic origin of each sequence was
416 determined based on individuals voluntarily disclosing their present place of residence, which can be
417 prone to inaccuracies. Likewise, missing location metadata meant we had to reduce the number of
418 sequences from each lineage to only sequences with known division when performing the DTA-GLM
419 phylogeographic analyses, which may have reduced the possible statistical power of the analyses.

420

421 **5. Conclusions**

422

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

430

431 **List of abbreviations**

432 **BF:** Bayes Factor

433 **COVID-19:** Coronavirus disease 2019

434 **Ct:** cycle threshold

435 **DTA:** discrete trait analysis

436 **ESS:** Effective Sample Size

437 **GLMs:** generalised-linear models

438 **HICs:** high income countries

439 **LDCs:** least developed countries

440 **LMICs:** low/middle-income countries

441 **MCC tree:** maximum clade credibility tree.
442 **MCMC:** Markov chain Monte Carlo
443 **rRT-PCR:** real-time reverse transcription polymerase chain reaction
444 **SARS-CoV-2:** severe acute respiratory syndrome coronavirus 2
445 **TMRCA:** time to the most recent common ancestor
446 **UN:** United Nations

447

448 **Declarations**

449 **Ethics approval and consent to participate:**

450 Patient consent for publication: Not applicable.

451 Ethics approval: The study was carried out under the ethical approval from the ethical approval
452 committee of Bangladesh Livestock Research Institute, Bangladesh with the reference number
453 BLRI/EA/2020102/2022, and with informed consent from tested individuals.

454

455 **Consent for publication:**

456 Not applicable.

457

458 **Availability of data and materials:**

459 The datasets and BEAST XML files supporting the conclusions of this article are available in the
460 ‘SARS_CoV_2_Bangladesh’ repository, [https://github.com/lorcancarnegie/SARS_CoV_2_Bangladesh].

461 The dataset for the newly reported genetic sequence data supporting the conclusions of this article are
462 available in GISAID EpiFlu database (<https://gisaid.org>) under the accession numbers included within
463 the additional files of this article [Table S6].

464

465 **Competing interests:**

466 We declare that we have no competing interests.

467

468 **Funding:**

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

480

481 **Authors' contributions:**

482 L.C., S.C.H, G.F. and J.R. conceived the phylodynamic and epidemiological analysis. L.C. wrote the
483 manuscript and performed the phylodynamic and epidemiological analysis. S.C.H, J.R. and G.F. assisted
484 with phylodynamic and epidemiological analysis. JT.M. and L.D.P. performed the viral lineage analyses.
485 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.
486 contributed to the collection of epidemiological data, laboratory testing, and generation of sequences
487 and helped with interpreting the findings. All other authors contributed to fieldwork or laboratory work
488 design or implementation. All authors gave final approval for publication.

489

490 **Acknowledgements:**

491 We thank the Child Health Research Foundation (CHRF), Bangladesh, for allowing us to use their data
492 on the GISAID EpiCoV database (accession numbers in **Table S7-8**). We also acknowledge the use of

493 additional sequences from the GISAID EpiCoV database for the global dataset used in our lineage
494 analyses and provide accession numbers and acknowledgements in the supplementary electronic
495 material (Table S7-8). We thank the Directorate General of Health Services (DGHS) and Institute of
496 Epidemiology Disease Control and Research (IEDCR) Bangladesh, DGHS, Ministry of Health and Family
497 Welfare for sharing COVID-19 samples and data.

498

499 Footnotes:

500 S.C.H and M.S are joint last authors.

501 The findings and conclusions in this report are those of the author(s) and do not necessarily represent
502 the official position of the U.S. Centers for Disease Control and Prevention.

503

504 References

- 505 [1] Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, et al. Genomic Epidemiology of
506 SARS-CoV-2 in Guangdong Province, China. *Cell* 2020;181:997-1003.e9.
507 <https://doi.org/10.1016/j.cell.2020.04.023>.
- 508 [2] Louis du P, T. MJ, E. ZA, Verity H, Christopher R, Bernardo G, et al. Establishment
509 and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* (1979)
510 2021;371:708–12. <https://doi.org/10.1126/science.abf2946>.
- 511 [3] Featherstone LA, Di Giallonardo F, Holmes EC, Vaughan TG, Duchêne S. Infectious
512 disease phylodynamics with occurrence data. *Methods in Ecology and Evolution*
513 2021;12:1498–507. <https://doi.org/https://doi.org/10.1111/2041-210X.13620>.
- 514 [4] Raghwani J, du Plessis L, McCrone JT, Hill SC, Parag K V, Thézé J, et al. Genomic
515 Epidemiology of Early SARS-CoV-2 Transmission Dynamics, Gujarat, India.
516 *Emerging Infectious Diseases* 2022;28:751–8. <https://doi.org/10.3201/eid2804.212053>.
- 517 [5] Giovanetti M, Slavov SN, Fonseca V, Wilkinson E, Tegally H, Patané JSL, et al.
518 Genomic epidemiology of the SARS-CoV-2 epidemic in Brazil. *Nature Microbiology*
519 2022;7:1490–500. <https://doi.org/10.1038/s41564-022-01191-z>.
- 520 [6] Tegally H, Wilkinson E, Lessells RJ, Giandhari J, Pillay S, Msomi N, et al. Sixteen
521 novel lineages of SARS-CoV-2 in South Africa. *Nature Medicine* 2021;27:440–6.
522 <https://doi.org/10.1038/s41591-021-01255-3>.
- 523 [7] Coppée R, Blanquart F, Jary A, Leducq V, Ferré VM, Franco Yusti AM, et al.
524 Phylodynamics of SARS-CoV-2 transmissions in France, Europe and the world during
525 2020 2022. <https://doi.org/10.1101/2022.08.10.22278636>.
- 526 [8] Levin AT, Owusu-Boaitey N, Pugh S, Fosdick BK, Zwi AB, Malani A, et al.
527 Assessing the burden of COVID-19 in developing countries: systematic review, meta-
528 analysis and public policy implications. *BMJ Global Health* 2022;7:e008477.
529 <https://doi.org/10.1136/bmjgh-2022-008477>.

- 530 [9] Usuzaki T, Chiba S, Shimoyama M, Ishikuro M, Obara T. A disparity in the number of
531 studies related to COVID-19 and SARS-CoV-2 between low- and middle-income
532 countries and high-income countries. *International Health* 2021;13:379–81.
533 <https://doi.org/10.1093/inthealth/ihaa088>.
- 534 [10] Nations U. UN: List of LDCs n.d. <https://www.un.org/ohrlls/content/list-ldcs> (accessed
535 February 7, 2024).
- 536 [11] Nations U. UN: LDC Identification Criteria & Indicators n.d.
537 [https://www.un.org/development/desa/dpad/least-developed-country-category/ldc-
538 criteria.html](https://www.un.org/development/desa/dpad/least-developed-country-category/ldc-
538 criteria.html) (accessed February 7, 2024).
- 539 [12] Brito A, Semenova E, Dudas G, Hassler G, Kalinich C, Kraemer M, et al. Global
540 disparities in SARS-CoV-2 genomic surveillance. *Nature Communications*
541 2022;13:7003. <https://doi.org/10.1038/s41467-022-33713-y>.
- 542 [13] Hodgins S, Saad A. Will the Higher-Income Country Blueprint for COVID-19 Work
543 in Low- and Lower Middle-Income Countries? *Global Health, Science and Practice*
544 2020;8:136–43. <https://doi.org/10.9745/GHSP-D-20-00217>.
- 545 [14] Cowley LA, Afrad MH, Rahman SIA, Mamun MM Al, Chin T, Mahmud A, et al.
546 Genomics, social media and mobile phone data enable mapping of SARS-CoV-2
547 lineages to inform health policy in Bangladesh. *Nature Microbiology* 2021;6:1271–8.
548 <https://doi.org/10.1038/s41564-021-00955-3>.
- 549 [15] Laxminarayan R, Wahl B, Dudala SR, Gopal K, Mohan B C, Neelima S, et al.
550 Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science*
551 (New York, NY) 2020;370:691–7. <https://doi.org/10.1126/science.abd7672>.
- 552 [16] Shishir TA, Naser I Bin, Faruque SM. In silico comparative genomics of SARS-CoV-2
553 to determine the source and diversity of the pathogen in Bangladesh. *PLOS ONE*
554 2021;16:e0245584.
- 555 [17] Sharif N, Dey SK. Phylogenetic and whole genome analysis of first seven SARS-CoV-
556 2 isolates in Bangladesh. *Future Virology* 2020;15:735–46. [https://doi.org/10.2217/fvl-
2020-0201](https://doi.org/10.2217/fvl-
557 2020-0201).
- 558 [18] Parvez MSA, Rahman MM, Morshed MN, Rahman D, Anwar S, Hosen MJ. Genetic
559 analysis of SARS-CoV-2 isolates collected from Bangladesh: Insights into the origin,
560 mutational spectrum and possible pathomechanism. *Computational Biology and*
561 *Chemistry* 2021;90:107413. <https://doi.org/10.1016/j.compbiolchem.2020.107413>.
- 562 [19] Anwar S, Nasrullah M, Hosen MJ. COVID-19 and Bangladesh: Challenges and How
563 to Address Them. *Frontiers in Public Health* 2020;8:154.
564 <https://doi.org/10.3389/fpubh.2020.00154>.
- 565 [20] Sharif N, Dey SK. Impact of population density and weather on COVID-19 pandemic
566 and SARS-CoV-2 mutation frequency in Bangladesh. *Epidemiology and Infection*
567 2021;149:e16. <https://doi.org/10.1017/S0950268821000029>.
- 568 [21] Streatfield PK, Karar ZA. Population challenges for Bangladesh in the coming
569 decades. *Journal of Health, Population, and Nutrition* 2008;26:261–72.
- 570 [22] Khan MR, Huq S, Risha AN, Alam SS. High-density population and displacement in
571 Bangladesh. *Science* 2021;372:1290–3. <https://doi.org/10.1126/science.abi6364>.
- 572 [23] Ahammad I, Hossain MU, Rahman A, Chowdhury ZM, Bhattacharjee A, Das KC, et
573 al. Wave-wise comparative genomic study for revealing the complete scenario and
574 dynamic nature of COVID-19 pandemic in Bangladesh. *PLOS ONE*
575 2021;16:e0258019.
- 576 [24] Saha O, Islam I, Shatadru RN, Rakhi NN, Hossain MdS, Rahaman MdM. Temporal
577 landscape of mutational frequencies in SARS-CoV-2 genomes of Bangladesh: possible
578 implications from the ongoing outbreak in Bangladesh. *Virus Genes* 2021;57:413–25.
579 <https://doi.org/10.1007/s11262-021-01860-x>.

- 580 [25] Islam A, Sayeed MA, Rahman MK, Zamil S, Abedin J, Saha O, et al. Assessment of
581 basic reproduction number ($R(0)$), spatial and temporal epidemiological determinants,
582 and genetic characterization of SARS-CoV-2 in Bangladesh. *Infection, Genetics and*
583 *Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in*
584 *Infectious Diseases* 2021;92:104884. <https://doi.org/10.1016/j.meegid.2021.104884>.
- 585 [26] Hossain MA, Jahid MIK, Hossain KMA, Walton LM, Uddin Z, Haque MO, et al.
586 Knowledge, attitudes, and fear of COVID-19 during the Rapid Rise Period in
587 Bangladesh. *PloS One* 2020;15:e0239646.
588 <https://doi.org/10.1371/journal.pone.0239646>.
- 589 [27] Shammi M, Bodrud-Doza M, Islam ARMT, Rahman MM. Strategic assessment of
590 COVID-19 pandemic in Bangladesh: comparative lockdown scenario analysis, public
591 perception, and management for sustainability. *Environment, Development and*
592 *Sustainability* 2021;23:6148–91. <https://doi.org/10.1007/s10668-020-00867-y>.
- 593 [28] Cheah ESG, Malkin J, Free RC, Lee S-M, Perera N, Woltmann G, et al. A two-tube
594 combined TaqMan/SYBR Green assay to identify mycobacteria and detect single
595 global lineage-defining polymorphisms in *Mycobacterium tuberculosis*. *The Journal of*
596 *Molecular Diagnostics : JMD* 2010;12:250–6.
597 <https://doi.org/10.2353/jmoldx.2010.090030>.
- 598 [29] Witham PK, Yamashiro CT, Livak KJ, Batt CA. A PCR-based assay for the detection
599 of *Escherichia coli* Shiga-like toxin genes in ground beef. *Applied and Environmental*
600 *Microbiology* 1996;62:1347–53. <https://doi.org/10.1128/aem.62.4.1347-1353.1996>.
- 601 [30] Dip SD, Sarkar SL, Setu MdAA, Das PK, Pramanik MdHA, Alam ASMRU, et al.
602 Evaluation of RT-PCR assays for detection of SARS-CoV-2 variants of concern.
603 *Scientific Reports* 2023;13:2342. <https://doi.org/10.1038/s41598-023-28275-y>.
- 604 [31] Lu Y, Li L, Ren S, Liu X, Zhang L, Li W, et al. Comparison of the diagnostic efficacy
605 between two PCR test kits for SARS-CoV-2 nucleic acid detection. *Journal of Clinical*
606 *Laboratory Analysis* 2020;34:e23554. <https://doi.org/10.1002/jcla.23554>.
- 607 [32] Wen D, Yang S, Li G, Xuan Q, Guo W, Wu W. Sample-to-Answer and Routine Real-
608 Time RT-PCR: A Comparison of Different Platforms for SARS-CoV-2 Detection. *The*
609 *Journal of Molecular Diagnostics : JMD* 2021;23:665–70.
610 <https://doi.org/10.1016/j.jmoldx.2021.02.010>.
- 611 [33] Tyson JR, James P, Stoddart D, Sparks N, Wickenhagen A, Hall G, et al.
612 Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome
613 sequencing using nanopore. *BioRxiv : The Preprint Server for Biology* 2020.
614 <https://doi.org/10.1101/2020.09.04.283077>.
- 615 [34] Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from
616 vision to reality. *Euro Surveill* 2017;22. [https://doi.org/10.2807/1560-](https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494)
617 [7917.ES.2017.22.13.30494](https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494).
- 618 [35] O’Toole, A. , Hill V. Llama Software: local lineage and monophyly assessment v0.1
619 2020.
- 620 [36] Larsson A. AliView: a fast and lightweight alignment viewer and editor for large
621 datasets. *Bioinformatics* 2014;30:3276–8.
622 <https://doi.org/10.1093/bioinformatics/btu531>.
- 623 [37] COVID-19 Dynamic Dashboard for Bangladesh. Institute of Epidemiology Disease
624 Control and Research n.d. <http://103.247.238.92/webportal/pages/covid19.php#>.
- 625 [38] Parag K V. Improved estimation of time-varying reproduction numbers at low case
626 incidence and between epidemic waves. *PLoS Computational Biology*
627 2021;17:e1009347. <https://doi.org/10.1371/journal.pcbi.1009347>.

- 628 [39] Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M et al.
629 Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and
630 healthcare demand. 2020.
- 631 [40] Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic
632 analysis. *Virus Evolution* 2018;4:vex042. <https://doi.org/10.1093/ve/vex042>.
- 633 [41] du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, et al.
634 Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*
635 2021;371:708–12. <https://doi.org/10.1126/science.abf2946>.
- 636 [42] Gutierrez B, Márquez S, Prado-Vivar B, Becerra-Wong M, Guadalupe JJ, Candido
637 DDS, et al. Genomic epidemiology of SARS-CoV-2 transmission lineages in Ecuador.
638 *Virus Evolution* 2021;7:veab051. <https://doi.org/10.1093/ve/veab051>.
- 639 [43] Marco A. R. Ferreira MAS. Bayesian analysis of elapsed times in continuous-time
640 Markov chains. *Canadian Journal of Statistics* 2008;36:355-368.
- 641 [44] Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving
642 Bayesian population dynamics inference: a coalescent-based model for multiple loci.
643 *Mol Biol Evol* 2013;30:713–24. <https://doi.org/10.1093/molbev/mss265>.
- 644 [45] Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
645 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*
646 2015;32:268–74. <https://doi.org/10.1093/molbev/msu300>.
- 647 [46] Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in
648 Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* 2018;67:901–4.
649 <https://doi.org/10.1093/sysbio/syy032>.
- 650 [47] Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling
651 trees. *BMC Evol Biol* 2007;7:214. <https://doi.org/10.1186/1471-2148-7-214>.
- 652 [48] Lemey P, Minin VN, Bielejec F, Kosakovsky Pond SL, Suchard MA. A counting
653 renaissance: combining stochastic mapping and empirical Bayes to quickly detect
654 amino acid sites under positive selection. *Bioinformatics (Oxford, England)*
655 2012;28:3248–56. <https://doi.org/10.1093/bioinformatics/bts580>.
- 656 [49] RStudio Team. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA
657 2020.
- 658 [50] Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele
659 G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evolution*
660 2020;6:veaa061. <https://doi.org/10.1093/ve/veaa061>.
- 661 [51] Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of
662 heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*
663 2016;2:vew007–vew007. <https://doi.org/10.1093/ve/vew007>.
- 664 [52] Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for
665 the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* 2006;23:7–9.
666 <https://doi.org/10.1093/molbev/msj021>.
- 667 [53] Ghafari M, du Plessis L, Raghwan J, Bhatt S, Xu B, Pybus OG, et al. Purifying
668 Selection Determines the Short-Term Time Dependency of Evolutionary Rates in
669 SARS-CoV-2 and pH1N1 Influenza. *Molecular Biology and Evolution*
670 2022;39:msac009. <https://doi.org/10.1093/molbev/msac009>.
- 671 [54] Stefan AM, Gronau QF, Schönbrodt FD, Wagenmakers E-J. A tutorial on Bayes
672 Factor Design Analysis using an informed prior. *Behavior Research Methods*
673 2019;51:1042–58. <https://doi.org/10.3758/s13428-018-01189-8>.
- 674 [55] Rahman FN, Rahman AKMF, Iwuagwu AO, Dalal K. COVID-19 Transmission due to
675 Mass Mobility Before and After the Largest Festival in Bangladesh: An
676 Epidemiologic Study. *Inquiry : A Journal of Medical Care Organization, Provision and*
677 *Financing* 2021;58:469580211023464. <https://doi.org/10.1177/00469580211023464>.

- 678 [56] The New York Times. The Trouble With the Case Curve During the Holidays n.d.
679 <https://www.nytimes.com/interactive/2021/11/22/us/covid-data-holiday-averages.html>.
- 680 [57] Brum E, Saha S, Sania A, Tanmoy AM, Hooda Y, Tanni AA, et al. Surging COVID-
681 19 in Bangladesh driven by B.1.351 variant. *MedRxiv* 2021:2021.04.27.21255706.
682 <https://doi.org/10.1101/2021.04.27.21255706>.
- 683 [58] Gu H, Xie R, Adam DC, Tsui JL-H, Chu DK, Chang LDJ, et al. Genomic
684 epidemiology of SARS-CoV-2 under an elimination strategy in Hong Kong. *Nature*
685 *Communications* 2022;13:736. <https://doi.org/10.1038/s41467-022-28420-7>.
- 686 [59] Geoghegan JL, Ren X, Storey M, Hadfield J, Jelley L, Jefferies S, et al. Genomic
687 epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa
688 New Zealand. *Nature Communications* 2020;11:6351. <https://doi.org/10.1038/s41467-020-20235-8>.
- 690 [60] Islam MT, Talukder AK, Siddiqui MN, Islam T. Tackling the COVID-19 pandemic:
691 The Bangladesh perspective. *Journal of Public Health Research* 2020;9:1794.
692 <https://doi.org/10.4081/jphr.2020.1794>.
- 693 [61] Di Giallonardo F, Duchene S, Puglia I, Curini V, Profeta F, Cammà C, et al. Genomic
694 Epidemiology of the First Wave of SARS-CoV-2 in Italy. *Viruses* 2020;12.
695 <https://doi.org/10.3390/v12121438>.
- 696 [62] Islam J, Guo X, Ali MA, Islam MA, Qi X, Zhuang G. Spatial pattern of COVID-19 in
697 Bangladesh: an ecological study. *BMJ Open* 2021;11:e047566.
698 <https://doi.org/10.1136/bmjopen-2020-047566>.
- 699 [63] Rabbi MFA, Khan MdI, Hasan S, Chalita M, Hasan KN, Sufian A, et al. Large scale
700 genomic and evolutionary study reveals SARS-CoV-2 virus isolates from Bangladesh
701 strongly correlate with European origin and not with China. *BioRxiv*
702 2021:2021.01.17.425424. <https://doi.org/10.1101/2021.01.17.425424>.
- 703 [64] Rashid SR. Remittance practices in rural Bangladesh: A gendered analysis. *Migration*
704 *and Development* 2022;11:21–39. <https://doi.org/10.1080/21632324.2019.1697494>.
- 705 [65] IOM. Bangladesh: POINT OF ENTRY ANALYSIS COVID-19. 2021.
- 706 [66] Sayeed MdA, Ferdous J, Saha O, Islam S, Choudhury SD, Abedin J, et al.
707 Transmission Dynamics and Genomic Epidemiology of Emerging Variants of SARS-
708 CoV-2 in Bangladesh. *Tropical Medicine and Infectious Disease* 2022;7.
709 <https://doi.org/10.3390/tropicalmed7080197>.
- 710 [67] Hasan MS, Islam MT, Alam ASMRU, Sarkar SL, Rahman MS, Islam OK, et al. Initial
711 reports of the SARS-CoV-2 Delta variant (B.1.617.2 lineage) in Bangladeshi patients:
712 Risks of cross-border transmission from India. *Health Science Reports* 2021;4:e366.
713 <https://doi.org/10.1002/hsr2.366>.
- 714 [68] Sawadogo Y, Galal L, Belarbi E, Zongo A, Schubert G, Leendertz F, et al. Genomic
715 Epidemiology of SARS-CoV-2 in Western Burkina Faso, West Africa. *Viruses*
716 2022;14. <https://doi.org/10.3390/v14122788>.
- 717 [69] Butera Y, Mukantwari E, Artesi M, Umuringa J d'arc, O'Toole ÁN, Hill V, et al.
718 Genomic sequencing of SARS-CoV-2 in Rwanda reveals the importance of incoming
719 travelers on lineage diversity. *Nature Communications* 2021;12:5705.
720 <https://doi.org/10.1038/s41467-021-25985-7>.
- 721 [70] Lee JN, Mahmud M, Morduch J, Ravindran S, Shonchoy AS. Migration, externalities,
722 and the diffusion of COVID-19 in South Asia. *Journal of Public Economics*
723 2021;193:104312. <https://doi.org/https://doi.org/10.1016/j.jpubeco.2020.104312>.
- 724 [71] Rahaman KR, Mahmud MS, Mallick B. Challenges of Testing COVID-19 Cases in
725 Bangladesh. *International Journal of Environmental Research and Public Health*
726 2020;17. <https://doi.org/10.3390/ijerph17186439>.

- 727 [72] Brown TS, Engø-Monsen K, Kiang M V, Mahmud AS, Maude RJ, Buckee CO. The
728 impact of mobility network properties on predicted epidemic dynamics in Dhaka and
729 Bangkok. *Epidemics* 2021;35:100441.
730 <https://doi.org/https://doi.org/10.1016/j.epidem.2021.100441>.
- 731 [73] Financial Express. Commuting to Dhaka -- the options n.d.
732 [https://thefinancialexpress.com.bd/views/opinions/commuting-to-dhaka-the-options-](https://thefinancialexpress.com.bd/views/opinions/commuting-to-dhaka-the-options-1581178251)
733 [1581178251](https://thefinancialexpress.com.bd/views/opinions/commuting-to-dhaka-the-options-1581178251).
- 734 [74] González-Leonardo M, Rowe F, Fresolone-Caparrós A. Rural revival? The rise in
735 internal migration to rural areas during the COVID-19 pandemic. Who moved and
736 Where? *Journal of Rural Studies* 2022;96:332–42.
737 <https://doi.org/10.1016/j.jrurstud.2022.11.006>.
- 738 [75] Miller D, Martin MA, Harel N, Tirosh O, Kustin T, Meir M, et al. Full genome viral
739 sequences inform patterns of SARS-CoV-2 spread into and within Israel. *Nature*
740 *Communications* 2020;11:5518. <https://doi.org/10.1038/s41467-020-19248-0>.
- 741 [76] Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al.
742 Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 2020;369:1255–60.
743 <https://doi.org/10.1126/science.abd2161>.
744