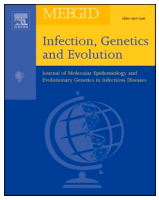
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Infection, Genetics and Evolution 95 (2021) 105057

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



Investigating the possible origin and transmission routes of SARS-CoV-2 genomes and variants of concern in Bangladesh

Abdullah Al Nahid, Ajit Ghosh *

Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh

ARTICLE INFO

Keywords:
SARS-CoV-2
COVID-19
Bangladesh
B.1.1.7
B.1.351
Epidemiology

ABSTRACT

The COVID-19 pandemic induced by the SARS-CoV-2 virus and its variants has ravaged most countries around the world including Bangladesh. We have analyzed publicly available genomic data to understand the current COVID-19 outbreak scenario as well as the evolutionary origin and transmission routes of SARS-CoV-2 isolates in Bangladesh. All the early isolates as well as recent B.1.1.7 and B.1.351 variants had already spread across the major divisional cities of Bangladesh. A sex biasness towards male COVID-19 patient samples sequencing has been observed over female patient samples in all age-group, that could be the trend in infection rate. Phylogenetic analysis indicated a total of 13 estimated countries, including Italy, India, United Kingdom, Saudi Arabia, United Arab Emirates, Germany, Australia, New Zealand, South Africa, Democratic Republic of the Congo, United States, Russia, and Denmark, could be the possible origin introduced SARS-CoV-2 isolates in Bangladesh because of regional and intercontinental travel. Recent, B.1.1.7 variant could be imported from a total of 7 estimated countries including UK, India, Nigeria, Spain, Ireland, Australia, and Indonesia, while South Africa and the United States are the most likely sources of B.1.351 variant in Bangladesh. Based on these findings, public health strategies could be designed and implemented to reduce the local transmission of the virus.

1. Introduction

The world has been going through a global pandemic of acute respiratory disease named the coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). SARS-CoV-2 is a novel strain of *betacoronavirus* within the Coronaviridae family (Gorbalenya et al., 2020), that was first reported from Wuhan, China, in December 2019 (Zhou et al., 2020). SARS-CoV-2 possesses an enveloped single-stranded, positive-sense ~30 kb long ribonucleic acid (RNA) genome (Naqvi et al., 2020). The longest segment of the genome at 5′ end encodes for orf1ab polyprotein while the rest of the genome consists of genes encoding four structural proteins including the spike glycoprotein (S), envelope protein (E), membrane protein (M), and nucleocapsid protein (N) (Hu et al., 2021; Michel et al., 2020). Since the identification of SARS-CoV-2, it has rapidly spread to 220 countries and territories including Bangladesh (https://www.worldometers.info/coronavirus), posing an unprecedented public health threat with over 165 million recorded cases of COVID-19, and more than 3.4 million deaths attributable to the disease worldwide as of May 20, 2021 (Dong et al., 2020). Bangladesh documented its first case of COVID-19 on March 8, 2020 (Imran et al., 2021). Since then, there have been 786,698 confirmed cases of the disease in the country, as well as 12,310 deaths with 1.56% case fatality rate (CFR) (Dong et al., 2020).

Over the devastating course of the SARS-CoV-2 pandemic, several variants have appeared in different regions of the world. Of them, a more infectious variant with D614G mutation in the spike glycoprotein quickly became dominant after emergence, owing to the increased human-to-human transmission efficiency, among the globally circulating strains during the early stages of the pandemic (Harrison et al., 2020; Korber et al., 2020; Volz et al., 2021a; Yurkovetskiy et al., 2020). According to a recent study, 98% of Bangladeshi SARS-CoV-2 isolates are of this common variant (Hasan et al., 2021). In addition to the pre-existing variants, several new variants of concern (VOCs) including B.1.1.7 or 501Y·V1, B.1.351 or 501Y·V2 and P.1 or 501Y·V3, have emerged independently from the United Kingdom (UK), South Africa, and Brazil respectively (Davies et al., 2021a; Faria et al., 2021; Tegally et al., 2021). Out of these three VOCs, presence of B.1.1.7 and B.1.351 variants had been reported in Bangladesh (https://virological.org/t/detection-of-the-b-1-1-7-and-b-1-351-sars-cov-2-variants-in-bangladesh/668), as well as in dozens of other countries around the globe as of April

* Corresponding author.
E-mail address: aghosh-bmb@sust.edu (A. Ghosh).

https://doi.org/10.1016/j.meegid.2021.105057
Received 25 May 2021; Received in revised form 15 August 2021; Accepted 30 August 2021
Available online 1 September 2021

9, 2021 (https://cov-lineages.org/global_report.html). Although much is still unknown about these VOCs, preliminary epidemiological and phylogenetic studies suggest the B.1.1.7 variant is 43 to 90% (95% CI 38–130) more transmissible than pre-existing variants (Davies et al., 2021a), with a 50 to 100% higher reproduction number (Volz et al., 2021b), and increased mortality (Davies et al., 2021b). Furthermore, findings from other preliminary studies indicate the B.1.351 variant is associated with higher viral load with potential advantage of enhanced transmissibility or immune escape (Cele et al., 2021; Tegally et al., 2021; Wibmer et al., 2021; Zhou et al., 2021). This variant may also amplify the risk of infection in people who have already been immunized (Planas et al., 2021).

As COVID-19 continues to wreak havoc due to rapid transformation of SARS-CoV-2, Bangladesh has begun a nationwide inoculation drive in order to restrain the virus, and administered more than 9.6 million doses of the Oxford-AstraZeneca vaccine till May 20, 2021 (https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/country_data/Bangladesh.csv). However, the daily COVID-19 cases and deaths in the country have seen a steep rise recently (Dong et al., 2020), which is found to be correlated with the increasing detection of B.1.351 variant circulating in the country (Saha et al., 2021). Another genomic variant surveillance in Bangladesh, conducted from January 1–March 24, 2021, observed a maximum 52% frequency of B.1.1.7 variant up until the second week of March 2021, and a dramatic rise in B.1.351 variant frequency of up to 81% in subsequent weeks (https://www.icddrb.org/news-and-events/news?id=874). In order to identify such emerging variants in the country, and monitor the viral evolution on a genomic level, scientists from Bangladesh have sequenced over 1500 SARS-CoV-2 genomes, and deposited the sequences in the Global Initiative on Sharing All Influenza Data (GISAID) database (Shu and McCauley, 2017). Analyses of these genomic data may aid in understanding the genetical and evolutionary features of the virus (Siddiqe and Ghosh, 2021), as well as in evaluating the outcome of various disease control strategies in Bangladesh, ranging from quarantine measures to travel restrictions, both locally and internationally, to reduce the transmission rate of evolving lineages.

An epidemiological analysis on genome samples from Bangladesh revealed regional sources of the isolates including variants rather than country-specific origin information (https://nextstrain.org/community/CHRF-Genomics/ncovBangladesh@main). Another phylogenetic study on 64 Bangladeshi genome samples reported Germany, United Kingdom, India, and Saudi Arabia as the major possible sources of the studied isolates with no report on the likely origin of the variants (Shishir et al., 2021). Moreover, Nextstrain (Hadfield et al., 2018) considered insufficient samples from Bangladesh for the global phylogenetics build of SARS-CoV-2 (https://nextstrain.org/ncov/global). Due to inadequate sampling of genome sequences from Bangladesh and lack of country-specific report on variant origin, most of these findings were not conclusive and fully reflective of the whole country. Thus, the recent rise in these VOCs necessitates a comprehensive genomic investigation of adequate and diverse samples to determine the country-specific transmission origin of Bangladeshi isolates including variants. As a result, this may help establish a firm understanding of viral epidemiology in the country, which is crucial for vaccine efficacy, clinical severity of infection, disease mortality, diagnostics and therapeutics, and public health safety precautions.

In this study, we aimed to identify the possible origin and transmission routes of SARS-CoV-2 isolates into Bangladesh, during the initial days of the pandemic, as well as for both B.1.1.7 and B.1.351 variants in recent times. We analyzed publicly available genomic data from countries of concern across all regions and estimated 3 separate time-resolved phylogenies with a total of 4157 subsampled genome sequences. By utilizing the genomic epidemiology approach to the data, we demonstrate that both the initial viral outbreak and the recent upsurge of VOCs in Bangladesh was the result of several independent introductions of the virus linked to international travel.

2. Materials and methods

2.1. Data acquisition and preprocessing

A total of 97,038 unique SARS-CoV-2 genome sequences and metadata were downloaded from the GISAID database(Shu and McCauley, 2017) and subsequently divided into 4 subsets (Table 1) for downstream analyses. As the genomic sample collection of this study was conducted on April 9, 2021, the G1 subset contained a total of 1009 complete, human-host genomes from Bangladesh (Dataset S1). From the subset G1, a set of 612 high-coverage genomes were considered for the initial phylogeny with a sample submission date till February 12, 2021, to exclude the VOC samples in Bangladesh, and the selected samples were added to the G2 subset (Dataset S2 and Table S1). In addition, 86,314 complete genomes from 48 additional countries and territories throughout the world with higher COVID-19 prevalence during the early stages of the outbreak were included to the G2 subset. A collection date prior to May 1, 2020, was applied to the foreign samples of G2 subset as a cutoff criterion to investigate the initial introductions of SARS-CoV-2 into Bangladesh which was first reported on March 8, 2020. For the G3 and G4 subsets, a total of 7507 and 3196 SARS-CoV-2 genomes, including both B.1.1.7 and B.1.351 variants of 20I/501Y·V1 and 20H/501Y·V2 clades, respectively, from 101 and 31 different countries and territories around the world with reported cases of the virus and studied variants, were selected for variant-focused phylogenies. 5100 of all G3 samples including 17 Bangladeshi samples (Dataset S3 and Table S2) belonged to the 20I/501Y·V1 clade, while 2255 of all G4 samples along with 47 samples from Bangladesh (Dataset S4 and Table S3) belonged to the 20H/501Y·V2 clade. Sample submission cutoff date like G1 subset had been applied to both G3 and G4 subsets. At the time of sample collection for this study, the Bangladeshi genome samples in both G3 and G4 subsets represented all the corresponding VOC samples of the country deposited to GISAID, whereas the foreign samples in both subsets represented diversified subsampled VOC and non-VOC genome sequences from regions with the criterion to properly construct the phylogenetic trees and estimate as many potential origins as possible for the genomes of interest. The caveat of such subsampling is further described in the discussion section. We are thankful to all the contributing laboratories and scientists for the viral sequences deposited to GISAID that have been used in this analysis.

Additionally, daily COVID-19 confirmed case related data of Bangladesh were downloaded from the World Health Organization (WHO) website (https://covid19.who.int/WHO-COVID-19-global-data.csv). Sample metadata of all subsets were checked in terms of data availability and any sample with incomplete collection date were excluded. Prior to phylogenetic analysis, G2, G3 and G4 subset samples were quality filtered and genome sequences with ambiguous characters were omitted.

Table 1
Summary of collected samples in subsets.

Subset	Total Countries and Territories	Sample Time Period	Total Samples	Bangladeshi Samples
G1	1	Submission till April 9, 2021	1009	1009
G2	49	**For Bangladeshi samples:** Submission till February 12, 2021 **For rest of the samples:** Collection till April 30, 2020	86,926	612
G3	101	Submission till April 9, 2021	7507	17
G4	31	Submission till April 9, 2021	3196	47

2.2. Exploratory data analysis

The metadata of G1 subset samples was analyzed with Python version 3.8.5 (Van Rossum and Drake, 2009) to investigate the Nextstrain clade distribution among divisions, as well as the gender and age-group distribution in Bangladeshi samples. G1 samples with no division data were labeled NA (Not Available), and the missing gender and age-group data were discarded before plotting with ggplot2 version 3.3.3 (Wickham, 2016) and tidyverse version 1.3.0 (Wickham et al., 2019) packages of R programming language version 3.6.3 (R Core Team, 2013).

2.3. Phylogenetic analysis

Quality filtered sequences of G2 subset were aligned to the reference SARS-CoV-2 sequence (NC_045512.2) (Wu et al., 2020) using MAFFT version 7.475 (Katoh and Standley, 2013). The aligned genomes were subsampled afterwards. All 612 genomes from Bangladesh within the G2 subset were assigned as focal samples during the subsampling process. In addition, 10 samples per month were chosen from 48 other countries and territories available in the subset based on genetic similarity to focal samples. A total of 1966 subsampled genomes were used to construct a maximum-likelihood phylogenetic tree with IQ-TREE version 2.0.3 (Minh et al., 2020) using the generalized time reversible (GTR) model. The tree was reconstructed with the inference of ancestral sequence states using Tree Time version 0.8.1 (Sagulenko et al., 2018) with the same GTR evolutionary model of nucleotide substitution. The complete phylogenetic analysis pipeline is available online (https://github.com/nextstrain/ncov, version 3). For tree visualizations, auspice version 0.5.0 (https://auspice.us) by Nextstrain (Hadfield et al., 2018) was used. Analyses performed on G2 subset samples were also performed on both G3 and G4 subsets to build phylogenetic trees for B.1.1.7 and B.1.351 variants using 1546 and 645 subsampled genomes, respectively. During the subsampling process of both subsets, 5 samples a month were selected for each country present in the corresponding subset except Bangladesh. The interactive phylogenetic builds are available online at: https://github.com/nahid18/ncov-bd.

2.4. Identification of possible origins of Bangladeshi genomes

To determine the potential ancestral origins of the Bangladeshi SARS-CoV-2 genomes, all three interactive phylogenetic builds were manually examined. The ancestral lineage is represented by the root of the tree, and the descendants of that ancestor are represented by the tips of the branches. Each branch point of the tree represents a divergence event where lies the most common ancestor of all the groups descended from that branch point. Genome sequences were considered to be more related to each other if they had a more recent ancestor. In the initial phylogeny, each branch tip carrying a Bangladeshi genome present in available Nextstrain clades were identified, and that branch was then inspected backwards in the tree until all the closely related exogenous strains and their country of origin were determined. In B.1.1.7 and B.1.351 variant-focused phylogenies, only the 20I/501Y·V1 and 20H/501Y·V2 Nextstrain clades were checked, respectively, for the origin prediction.

3. Results

3.1. Data exploration for the current scenario of COVID-19 cases in Bangladesh

To better understand the current COVID-19 situation in Bangladesh, the pattern of daily reported COVID-19 cases was analyzed. The highest peak of nearly 8000 confirmed cases per day was recently reported in the early April 2021 (Fig. 1A). The metadata of 1009 complete human-host SARS-CoV-2 genomes from Bangladesh (Dataset S1) accessible via GISAID (Shu and McCauley, 2017) were examined. The overall Nextstrain clade distribution across all the major divisional cities of Bangladesh (Fig. 1B), reveals that samples from Dhaka and Sylhet divisions contain the 20I/501Y·V1 clade (B.1.1.7 variant) and primarily start clustering around January 2021. Furthermore, samples from the Dhaka, Chattogram, Khulna, and Mymensingh divisions include the 20H/501Y·V2 clade (B.1.351 variant) that has been the most prevalent during March to April 2021. This raises the alarming possibility that these VOCs are now dispersed in Bangladesh. We have also looked at the gender and age-group distribution within the available samples and found out that samples from male patients are more frequently sequenced in almost every age-group except 10–19, than female patients

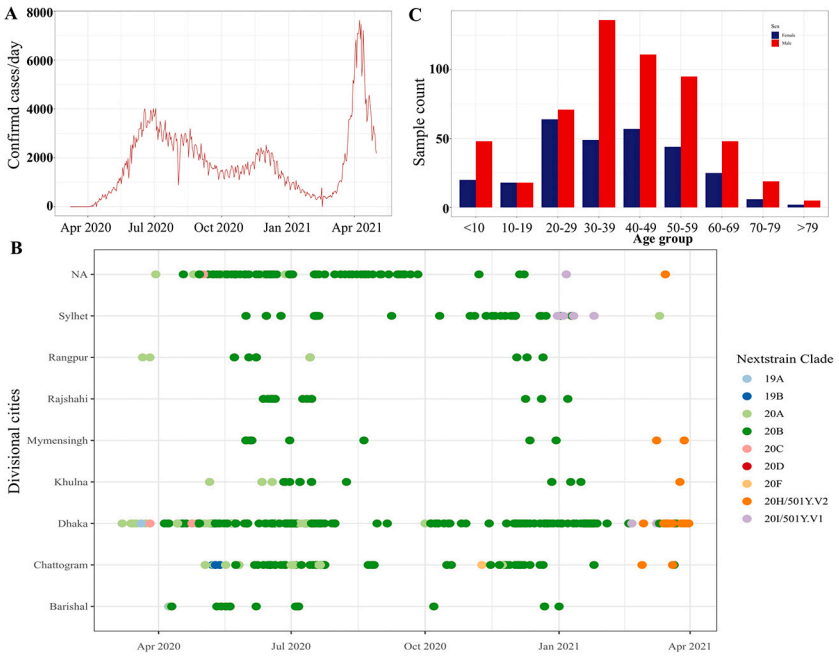


Fig. 1. Status of COVID-19 cases in Bangladesh. (A) Trend of COVID-19 confirmed cases in Bangladesh over the course of the pandemic. (B) Distribution of Nextstrain clades and Bangladeshi Divisions in SARS-CoV-2 samples isolated in Bangladesh (n = 1009). NA represents the samples with unassigned division data. (C) Distribution of sex and age-group in Bangladeshi SARS-CoV-2 samples (n = 836). The male and female samples are represented via red and dark-blue colors, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Fig. 1C and Table S4).

3.2. Phylogenetic analysis of SARS-CoV-2 isolates of Bangladesh

An independent phylogenetic analysis has been performed using a curated collection of 86,926 SARS-CoV-2 genome sequences covering all regions and most countries/territories where the COVID-19 infection rate was high during the early stages of the pandemic to precisely trace the origin of initial SARS-CoV-2 isolates of Bangladesh. The generated phylogenetic tree containing a total of 1966 subsampled genomes, showed that all 612 high-coverage genome sequences from Bangladesh are scattered across the tree (Fig. 2), suggesting multiple introductions of the virus into the country, as well as grouped in several distinct clusters indicating community transmission. The two well-defined clusters of Bangladeshi samples within 20B clade contain a total of 482 descendants and share common ancestry with sequences isolated in Italy (EPI_ISL_469050, EPI_ISL_494761, EPI_ISL_494774). In addition, multiple instances of Bangladeshi samples with an estimated foreign origin have been found in 5 separate clades distributed within the tree (Table 2). In detail, Bangladeshi sample with GISAID accession no. EPI_ISL_458133 is close to the sample EPI_ISL_509705 from the United States within 19A clade. Two samples from Bangladesh, EPI_ISL_450339 and EPI_ISL_450341, have high similarity with the Australian EPI_ISL_410718 sample of 19B clade. Another Bangladeshi sample EPI_ISL_450345 within 19B clade is showing high similarity with an Indian sample EPI_ISL_450324.

In addition to the large clusters of Bangladeshi samples, several clusters of few samples are identified as well throughout the tree, indicating the local transmission on a smaller scale. Within the 20A clade, EPI_ISL_445244 and 3 other samples from Bangladesh are close to Italian EPI_ISL_412973 and EPI_ISL_451306 samples. A group of 10 Bangladeshi samples including EPI_ISL_774877, EPI_ISL_605887 and EPI_ISL_774896 within the same clade have a shared ancestry with the EPI_ISL_707964 sample isolated in Germany. Another group of 11 Bangladeshi clustered as 20A, including EPI_ISL_774900 and EPI_ISL_605910, are like the Russian EPI_ISL_450288 sample. Similarly, a total of 19 Bangladeshi 20A

Table 2
Estimated origins of initial representative samples of Bangladesh.

Nextstrain Clade	Bangladeshi Sample	Estimated Origin
19A	EPI_ISL_458133	United States
19B	EPI_ISL_450339	Australia
	EPI_ISL_450341	
	EPI_ISL_450345	India
20A	EPI_ISL_477134	
	EPI_ISL_445244	Italy
	EPI_ISL_774877	Germany
	EPI_ISL_605887	
	EPI_ISL_774896	
	EPI_ISL_774900	Russia
	EPI_ISL_605910	
	EPI_ISL_605889	South Africa
	EPI_ISL_468070	Democratic Republic of the Congo
	EPI_ISL_774893	
	EPI_ISL_466637	Saudi Arabia
	EPI_ISL_450343	
20B	EPI_ISL_477130	United Kingdom
	EPI_ISL_774890	
	EPI_ISL_774889	
	EPI_ISL_483630	New Zealand
	EPI_ISL_464162	
	EPI_ISL_605884	United Arab Emirates
	EPI_ISL_514238	
	EPI_ISL_774956	
20C	EPI_ISL_477128	Denmark

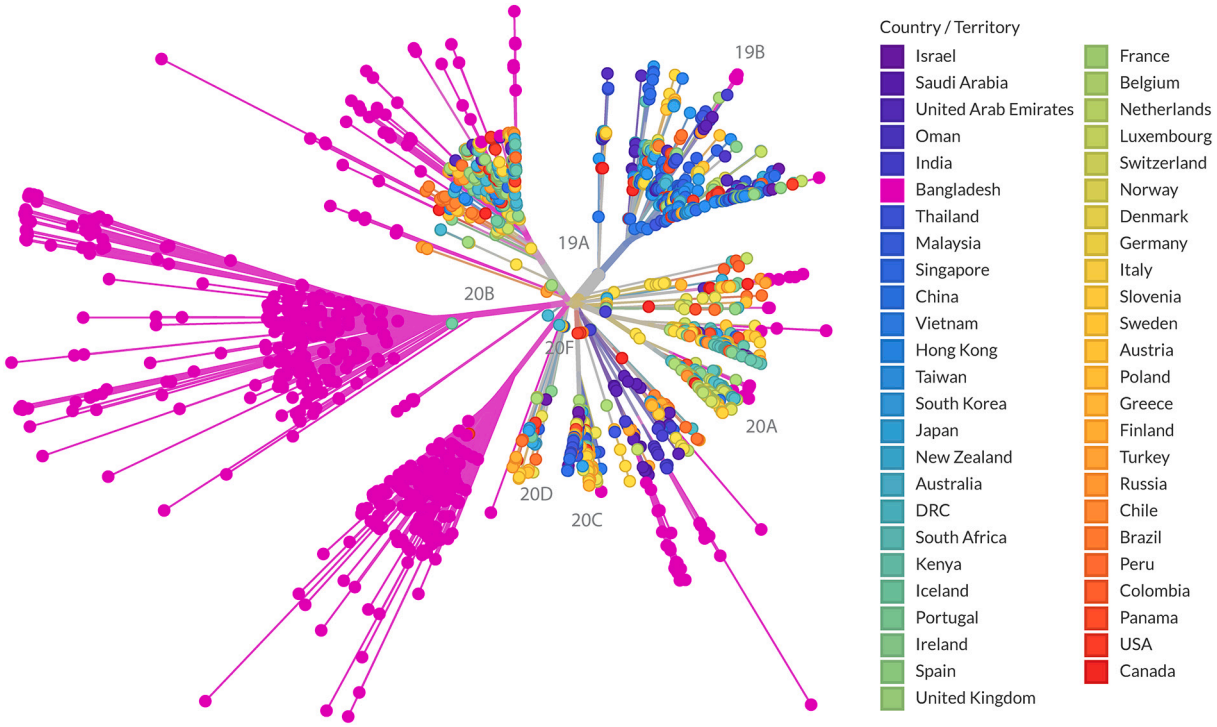


Fig. 2. Phylogenetic tree of initial SARS-CoV-2 isolates in an unrooted representation. Out of 86,926 analyzed genome sequences from 49 countries and territories including Bangladesh in the G2 genome subset, the tree was constructed using a total of 1966 subsampled complete genomes with IQ-TREE version 2.0.3 (Minh et al., 2020) and TreeTime version 0.8.1 (Sagulenko et al., 2018) utilizing the generalized time-reversible (GTR) model and visualized using auspice version 0.5.0 (https://auspice.us) by Nextstrain (Hadfield et al., 2018). Pink branch tips denote all 612 Bangladeshi genomes in the tree which are distinctively clustered as well as scattered throughout the phylogeny. The cutoff date criterion applied to the foreign samples are also visible in the tree. The interactive version of the tree can be accessed at: https://nextstrain.org/community/nahid18/ncov-bd@main/initial. We are thankful to all the authors who have generously deposited and shared the SARS-CoV-2 genome sequence data on GISAID. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

samples including EPI_ISL_466637 and EPI_ISL_450343 have close similarity to the EPI_ISL_513222 sample from Saudi Arabia. In addition, four samples within 20A clade from Bangladesh including EPI_ISL_468070 and EPI_ISL_774893 are highly related to the sample EPI_ISL_429255 from the Democratic Republic of the Congo (DRC). Indian EPI_ISL_435065, and South African EPI_ISL_450299 sample have close resemblance to the Bangladeshi EPI_ISL_477134 and EPI_ISL_605889 samples of the 20A clade, respectively.

Although most of the 20B samples from Bangladesh are found to have Italian ancestry, samples from the United Kingdom, United Arab Emirates and New Zealand are also showing high similarity to Bangladeshi clustered detected across the 20B clade. EPI_ISL_477130, EPI_ISL_774889, EPI_ISL_774890 and 5 other samples from Bangladesh are closely related to the EPI_ISL_488766 sample from the United Kingdom. A total of 10 samples from Bangladesh including EPI_ISL_483630 and EPI_ISL_464162 are close to the EPI_ISL_579166 sample isolated in New Zealand. Moreover, a combined set of 8 Bangladeshi samples of 20B clade in 3 separate groups are closely related to 3 samples with GISAID accession no. EPI_ISL_520738, EPI_ISL_520717 and EPI_ISL_528686 from the United Arab Emirates (UAE). In 20C clade, Bangladeshi EPI_ISL_477128 sample is highly similar to the EPI_ISL_614672 and EPI_ISL_614674 samples from Denmark.

Based on the evidence indicated by the phylogenetic tree, it is evident that SARS-CoV-2 isolates have been introduced into Bangladesh from several countries of all regions, except South America, during the initial outbreak of COVID-19. Multiple introductions from a single country such as Italy, India and UAE have occurred as well.

3.3. Phylogenetic analysis of B.1.1.7 variant samples in Bangladesh as compared to other countries

A time-resolved phylogenetic tree was created containing 1546 subsampled genomes including 641 samples of B.1.1.7 variant to develop a better understanding of the evolutionary relationship and potential propagation patterns of this variant in Bangladesh. All 17 Bangladeshi samples are found scattered across the 20I/501Y·V1 clade within the tree (Fig. 3), suggesting multiple viral introductions into Bangladesh from several countries (Table 3). The sample EPI_ISL_1360439 from Bangladesh is closely related to the sample EPI_ISL_756846 isolated in the United Kingdom. Another Bangladeshi

Table 3
Estimated origins of B.1.1.7 variant representative samples of Bangladesh.

Nextstrain Clade	Bangladeshi Sample	Estimated Origin
20I/501Y·V1	EPI_ISL_890237	United Kingdom
	EPI_ISL_1360439	
	EPI_ISL_906091	Spain
	EPI_ISL_1508943	Indonesia
	EPI_ISL_1508946	
	EPI_ISL_906098	Ireland
	EPI_ISL_1360445	Australia
	EPI_ISL_1498132	India
	EPI_ISL_1509000	Nigeria
	EPI_ISL_1360451	
	EPI_ISL_1360430	
	EPI_ISL_1360446	
	EPI_ISL_1508954	
	EPI_ISL_1508955	

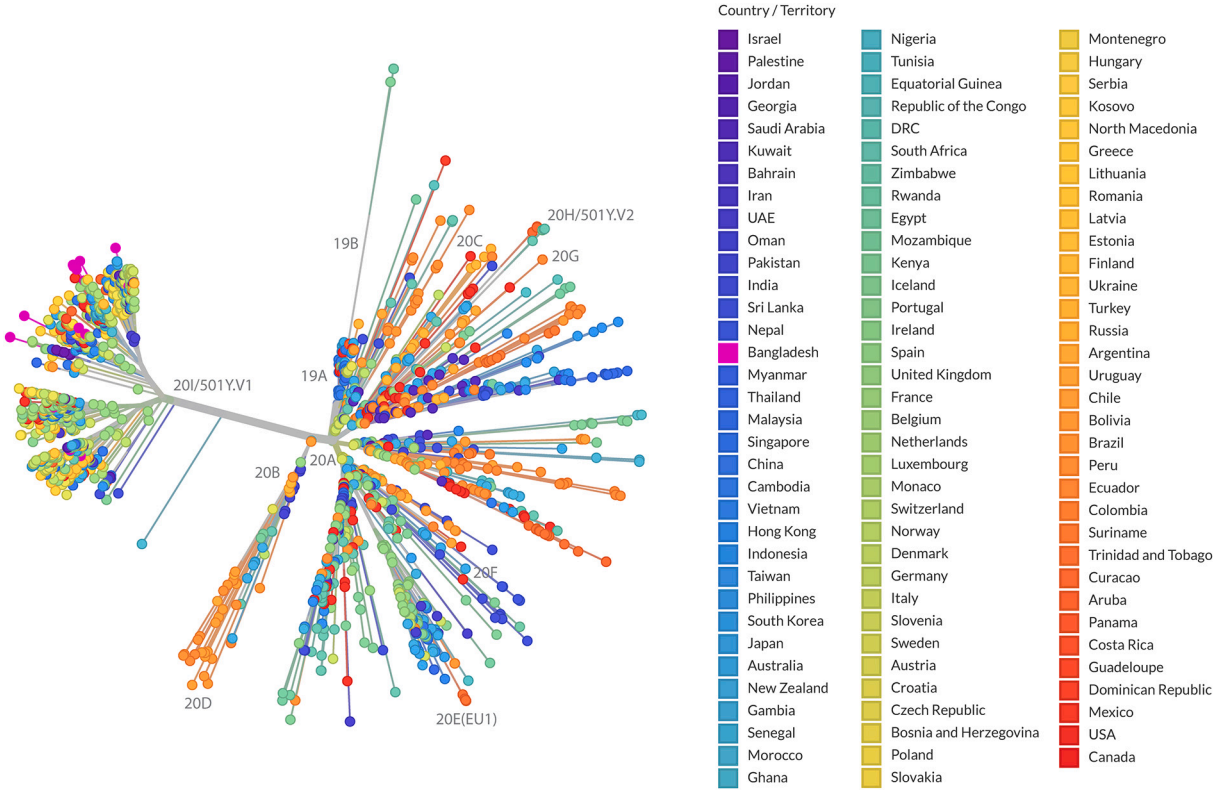


Fig. 3. B.1.1.7 variant focused construction of phylogenetic tree. Out of 7507 analyzed genomes in the G3 genome subset from 101 countries and territories including Bangladesh, the maximum-likelihood phylogenetic tree was built with a total of 1546 subsampled full-length genomes of which 641 genomes are of B.1.1.7 variant, using similar methods as Fig. 2 utilizing the GTR evolutionary model. The pink colored branch tips represent all 17 Bangladeshi B.1.1.7 variant samples which are visibly scattered across the tree. The tree can be interactively explored at: https://nextstrain.org/community/nahid18/ncov-bd@main/uk. We are thankful to all the authors who have generously deposited and shared the SARS-CoV-2 genome sequence data on GISAID. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sample EPI_ISL_906098 is highly similar to a sample EPI_ISL_837380 from Ireland. Two samples- EPI_ISL_1508943 and EPI_ISL_1508946, are close to the EPI_ISL_1118931 sample from Indonesia. A total of 6 samples including EPI_ISL_1509000 and EPI_ISL_1360451 are related to the Nigerian EPI_ISL_985080, EPI_ISL_872625 samples. Three distinct Bangladeshi samples (EPI_ISL_1498132, EPI_ISL_1360445 and EPI_ISL_906091) have shown similarity with samples from India (EPI_ISL_995709), Australia (EPI_ISL_812441) and Spain (EPI_ISL_1059975), respectively.

3.4. Phylogenetic analysis of B.1.351 variant Bangladeshi samples from global perspective

To identify the origin of the B.1.351 variant isolates in Bangladesh, we analyzed 3196 samples from 31 countries and territories, the bulk of which were B.1.351 variant. The resulting phylogenetic tree (Fig. 4) generated from the 645 subsampled sequences, reveals 46 out of 47 B.1.351 variant samples of Bangladesh clustered under a single node within 20H/501Y·V2 clade, manifesting clear signs of local transmission with single ancestral source, with paternal node being shared with samples isolated in South Africa. Samples from France and Turkey are placed close to the Bangladeshi samples in the phylogenetic tree. The remaining lone sample (EPI_ISL_1508892) that did not cluster with other samples had the closest ancestral genotype of United States origin (EPI_ISL_1097839). Based on the data (Table 4), it is highly likely that South Africa and the United States are the estimated origins to have introduced the B.1.351 variant to Bangladesh.

Table 4
Estimated origins of B.1.351 variant representative samples of Bangladesh.

Nextstrain Clade	Bangladeshi Sample	Estimated Origin
20H/501Y·V2	EPI_ISL_1508892	United States
	EPI_ISL_943561	South Africa
	EPI_ISL_1360428	
	EPI_ISL_1360447	
	EPI_ISL_1360448	
	EPI_ISL_1498150	

4. Discussions

As the global pandemic of COVID-19 progresses, new SARS-CoV-2 variants that are highly pathogenic and possibly more transmissible in nature than pre-existing variants continues to emerge due to mutational changes in the viral genome, causing a second wave of COVID-19 in several regions and countries around the world (Cacciapaglia et al., 2020; Salyer et al., 2021). From the COVID-19 confirmed cases data, a severe second wave is detected in Bangladesh from late March to early April 2021 (Fig. 1A) as compared to the first wave of April to August 2020 (Imran et al., 2021). The exploratory analysis on Bangladeshi genomic sequences reveals a high prevalence of B.1.351 variants could be a major determining factor for the second wave of COVID-19 in the country, which is consistent with an already published study (Saha et al., 2021). Moreover, we have reported that both B.1.1.7 and B.1.351 variant samples in Bangladesh are now dispersed in several divisions of Bangladesh including Dhaka, Sylhet, Chattogram, Mymensingh, and Khulna (Fig. 1B). A continuous trend of male-female disparity in terms of sample sequencing in almost every age group was observed (Fig. 1C), with more male patient samples being sequenced than female patient samples. This is most likely attributed to the fact that male COVID-19 patients has a far higher infection and mortality rate than females, as several studies have reported similar gender-based differentiation of COVID-19 severity in Bangladesh (Scully et al., 2020; Siam et al., 2020). Biological sex plays an important role in the COVID-19 infection

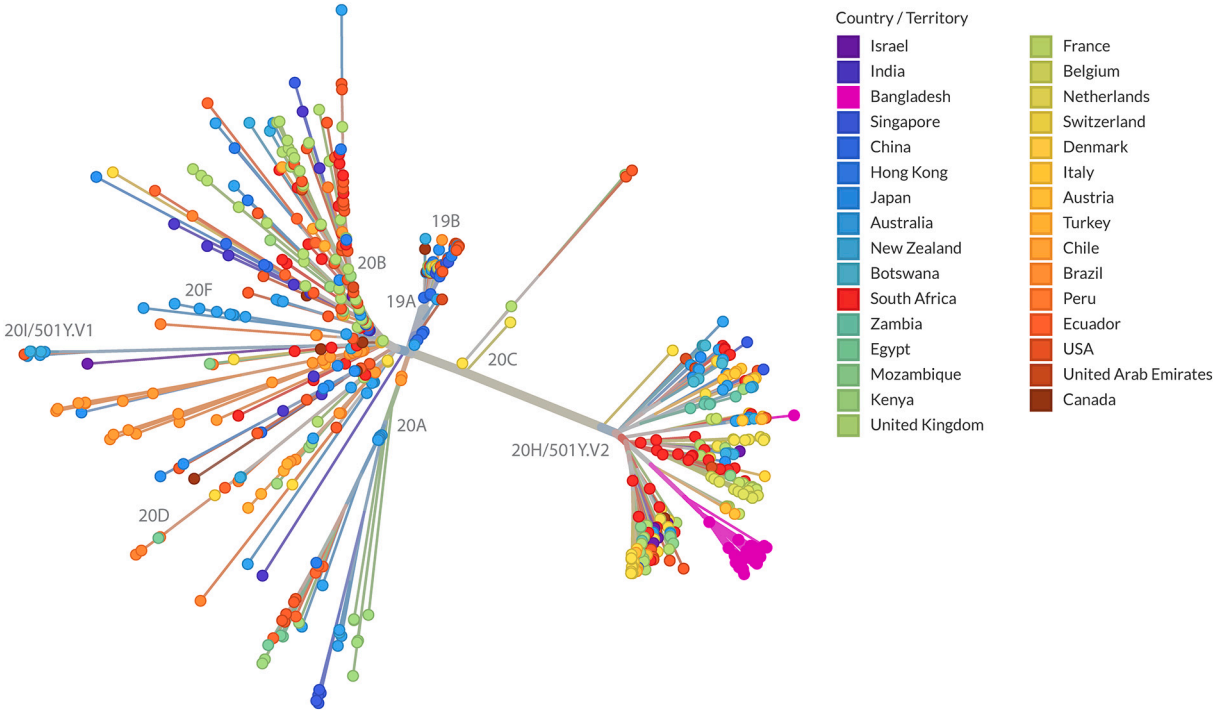


Fig. 4. Construction of B.1.351 variant focused phylogenetic tree. Out of 3196 analyzed samples from 31 countries and territories including Bangladesh in the G4 genome subset, the maximum-likelihood phylogenetic tree was constructed with a total of 645 subsampled complete genomes of which 256 are of B.1.351 variant, using similar methods as Fig. 2 and Fig. 3 utilizing the same GTR substitution model. Pink colored branch tips are the representatives of all 47 B.1.351 variant samples of Bangladesh, which are distinctively grouped under a common ancestor suggesting local transmission of the variant The tree is available for interactive inspection at: https://nextstrain.org/community/nahid18/ncov-bd@main/sa. We are thankful to all the authors who have generously deposited and shared the SARS-CoV-2 genome sequence data on GISAID. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

outcome by manifesting the infection susceptibility, pathogenesis, innate and adaptive immune responses, level of inflammation, and capacity of tissue repair. Presence of a more robust T cell activation mechanism and higher levels of innate immune cytokines in female as compared to male patients, could be the possible explanation for this kind of gender biasness in COVID-19 (Takahashi et al., 2020). A male dominance in COVID-19 mortality rate has been consistent with the other viral pathogenesis (Scully et al., 2020).

Phylogenetic analyses on all the subsets have helped us to unravel the estimated origins of SARS-CoV-2 isolates as well as both B.1.1.7 and B.1.351 variants in Bangladesh. Based on the evidence presented in the phylogenetic trees (Fig. 2-4), a total of 17 unique countries from all regions of the world, except South America, have collectively introduced initial SARS-CoV-2 isolates as well as two analyzed VOCs into Bangladesh (Fig. 5). Of these countries, initial Bangladeshi samples are showing similarity with the sequences reported from Italy, India, UK, Saudi Arabia, UAE, Germany, Australia, New Zealand, South Africa, DRC, USA, Russia, and Denmark. A subset of these countries was reported to be the estimated origins of SARS-CoV-2 in Bangladesh in few other studies (Parvez et al., 2021; Shishir et al., 2021). Multiple introductions from a single country including Italy, India, UK, Saudi Arabia have occurred as well. Furthermore, B.1.1.7 variant samples of Bangladesh are found to have close similarity to the samples from UK, India, Nigeria, Spain, Ireland, Australia, and Indonesia. In case of B.1.351 variant, it is highly possible that South Africa and the USA are the estimated origins of this variant in Bangladesh. It is evident from these findings that multiple variants from several countries including UK, India, South Africa, USA, and Australia have arrived in Bangladesh because of regional and intercontinental travel. Based on the widespread distribution of several distinct clusters throughout the phylogenetic tree, these foreign isolates might be transmitted locally in different areas of Bangladesh during the COVID-19 pandemic.

Findings from this study should be taken into consideration within the context of a few caveats. The foreign samples of G3 and G4 subsets contain a diverse selection but not all the samples of all reported countries. This remain a caveat of the analysis as it can possibly leave exogenous SARS-CoV-2 strains that can be close to the Bangladeshi genomes. Similarly, during the subsampling process of phylogenetic tree construction, a fixed number of samples were selected for the final grouping, hence may not include all foreign samples that are possibly close to the focal samples. However, this study rather remains a significant snapshot of the selected samples and from the phylogenetic trees it is evident that all the Bangladeshi genomes of interest either have an exogenous origin or are descendants of the local samples. As more variants continue to appear in different parts of the world due to the mutating abilities of SARS-CoV-2, the list of possible foreign origins of Bangladeshi genomes can increase rapidly. In addition to the subsampling limitations, samples selected for the variant phylogenies may be biased towards the country of initial emergence of the corresponding variants. Despite these limitations, the findings regarding the origin of these variants are very alarming and should be brought to the notice of policymakers and public health authorities right away to ensure the epidemiological surveillance of SARS-CoV-2 in Bangladesh on a much broader scale to prevent further escalation of viral transmission on a community level in the country.

5. Conclusions

This study illustrates the contribution of both regional and intercontinental travel in the spread of SARS-CoV-2 VOCs in Bangladesh. Continuous genomic surveillance is crucial while the COVID-19 pandemic unfolds to closely monitor the ever-changing circulation pattern of SARS-CoV-2 variants in Bangladesh since these findings have major ramifications for mitigation and vaccination tactics. Given how quickly new variants can propagate across the world due to air travel without being detected for a long time, these findings are critical for administrative authorities and public health officials in terms of implementing effective interventions and guidelines including but not limited to travel restrictions and mandatory quarantine or lockdown measurements to prevent further proliferation of these variants in Bangladesh. Immunization efforts must also be continued and enforced to keep the infection rate down.

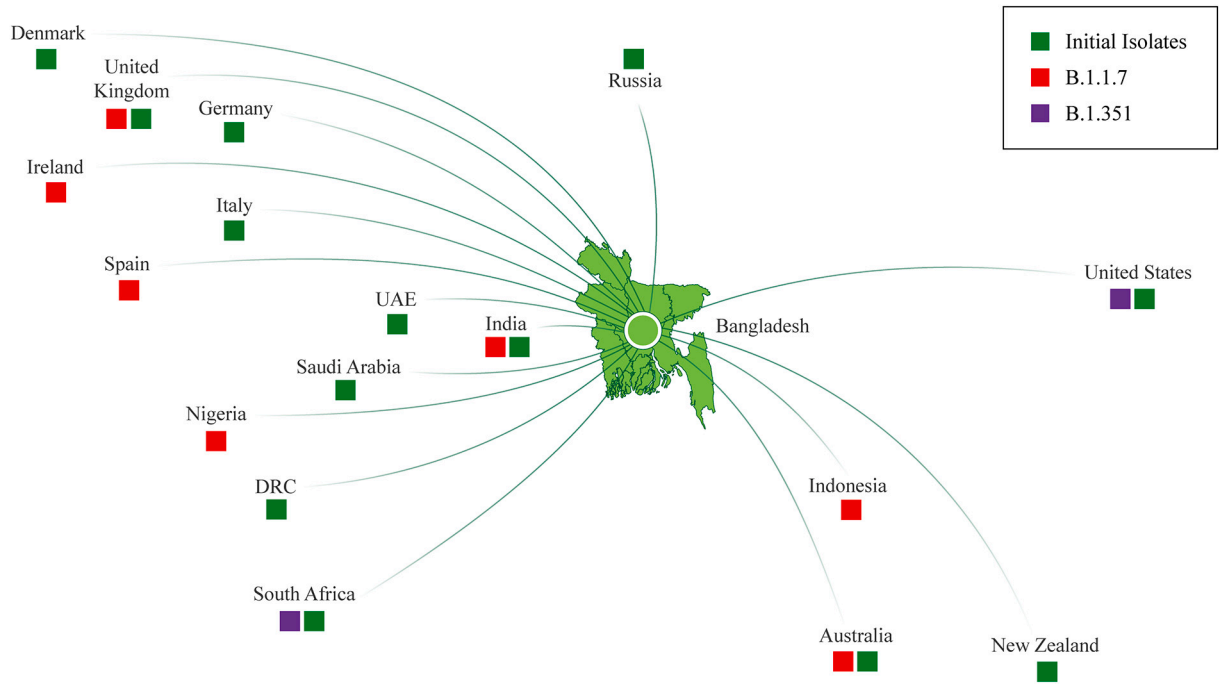


Fig. 5. Estimated origins and possible transmission routes of early SARS-CoV-2 isolates and B.1.1.7, B.1.351 variants into Bangladesh. Multiple countries from all regions except South America are found to have introduced SARS-CoV-2 into Bangladesh. India, United Kingdom, United States, Australia, and South Africa have been estimated to be the potential origins of multiple variants of the virus in Bangladesh.

Data availability

The authors declare that all the data will be available without any restrictions.

Funding

There was no funding for this study.

Authors' contributions

AG conceived the idea and designed the experiments. AAN performed all the experiments including data acquisition, analysis, and interpretation. AAN wrote the initial draft of the manuscript. Both authors read the manuscript and approved the final version.

CRediT authorship contribution statement

Abdullah Al Nahid: Data curation, Software, Investigation, Visualization, Writing – original draft. **Ajit Ghosh:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that there is no competing interest.

Acknowledgements

Authors gratefully acknowledge and thank all the contributing laboratories and scientists for the viral sequences deposited to GISAID that have been used in this study. Authors also acknowledge the logistic support and laboratory facilities of the Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet, Bangladesh.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.105057.

References

Cacciapaglia, G., Cot, C., Sannino, F., 2020. Second wave COVID-19 pandemics in Europe: a temporal playbook. Sci. Rep. 10, 15514. https://doi.org/10.1038/s41598-020-72611-5.

Cele, S., Gazy, I., Jackson, L., Hwa, S.-H., Tegally, H., Lustig, G., Giandhari, J., Pillay, S., Wilkinson, E., Naidoo, Y., Karim, F., Ganga, Y., Khan, K., Bernstein, M., Balazs, A.B., Gosnell, B.I., Hanekom, W., Moosa, M.-Y.S., Lessells, R.J., de Oliveira, T., Sigal, A., 2021. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. Nature 1–6. https://doi.org/10.1038/s41586-021-03471-w.

Davies, N.G., Jarvis, C.I., Edmunds, W.J., Jewell, N.P., Diaz-Ordaz, K., Keogh, R.H., 2021b. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature 1–5. https://doi.org/10.1038/s41586-021-03426-1.

Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A.B., Russell, T.W., Tully, D.C., Washburne, A.D., Wenseleers, T., Gimma, A., Waites, W., Wong, K.L.M., van Zandvoort, K., Silverman, J.D., Group, C. C.-19 W, Consortium, C.-G.U. (COG-U), Diaz-Ordaz, K., Keogh, R., Eggo, R.M., Funk, S., Jit, M., Atkins, K.E., Edmunds, W.J., 2021a. Estimated Transmissibility and Impact of SARS-CoV-2 Lineage B.1.1.7 in England. Science 372. https://doi.org/10.1126/science.abg3055.

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20, 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1.

Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D. da S., Mishra, S., Crispim, M.A.E., Sales, F.C.S., Hawryluk, I., McCrone, J.T., Hulswit, R.J.G., Franco, L.A.M., Ramundo, M.S., de Jesus, J.G., Andrade, P.S., Coletti, T.M., Ferreira, G.M., Silva, C.A.M., Manuli, E.R., Pereira, R.H.M., Peixoto, P.S., Kraemer, M.U.G., Gaburo, N., Camilo, C. da C., Hoeltgebaum, H., Souza, W.M., Rocha, E.C., de Souza, L.M., de Pinho, M.C., Araujo, L.J.T., Malta, F.S.V., de Lima, A. B., Silva, J. do P., Zauli, D.A.G., Ferreira, A.C. de S., Schnekenberg, R.P., Laydon, D. J., Walker, P.G.T., Schlüter, H.M., dos Santos, A.L.P., Vidal, M.S., Caro, V.S.D., Filho, R.M.F., dos Santos, H.M., Aguiar, R.S., Proença-Modena, J.L., Nelson, B., Hay, J.A., Monod, M., Miscouridou, X., Coupland, H., Sonabend, R., Vollmer, M., Gandy, A., Prete, C.A., Nascimento, V.H., Suchard, M.A., Bowden, T.A., Pond, S.L.K., Wu, C.-H., Ratmann, O., Ferguson, N.M., Dye, C., Loman, N.J., Lemey, P., Rambaut, A., Fraiji, N.A., Carvalho, M. do P.S.S., Pybus, O.G., Flaxman, S., Bhatt, S., Sabino, E.C., 2021. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science. https://doi.org/10.1126/science.abh2644.

Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W., Penzar, D., Perlman, S., Poon, L.L.M., Samborskiy, D.V., Sidorov, I.A., Sola, I., Ziebuhr, J., Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. 5, 536–544. https://doi.org/10.1038/s41564-020-0695-z.

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34, 4121–4123. https://doi.org/10.1093/bioinformatics/bty407.

Harrison, A.G., Lin, T., Wang, P., 2020. Mechanisms of SARS-CoV-2 transmission and pathogenesis. Trends Immunol. 41, 1100–1115. https://doi.org/10.1016/j.it.2020.10.004.

Hasan, Md.M., Das, R., Rasheduzzaman, Md, Hussain, Md.H., Muzahid, N.H., Salauddin, A., Rumi, M.H., Mahbubur Rashid, S.M., Siddiki, A.Z., Mannan, A., 2021. Global and local mutations in Bangladeshi SARS-CoV-2 genomes. Virus Res. 297, 198390. https://doi.org/10.1016/j.virusres.2021.198390.

Hu, B., Guo, H., Zhou, P., Shi, Z.-L., 2021. Characteristics of SARS-CoV-2 and COVID-19. Nat. Rev. Microbiol. 19, 141–154. https://doi.org/10.1038/s41579-020-00459-7.

Imran, Md.A., Noor, I.U., Ghosh, A., 2021. Impact of lockdown measures and meteorological parameters on the COVID-19 incidence and mortality rate in Bangladesh. Infect. Microbes Dis. 3, 41–48. https://doi.org/10.1097/IM9.0000000000000052.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Angyal, A., Brown, R.L., Carrilero, L., Green, L.R., Groves, D.C., Johnson, K.J., Keeley, A.J., Lindsey, B.B., Parsons, P.J., Raza, M., Rowland-Jones, S., Smith, N., Tucker, R.M., Wang, D., Wyles, M.D., McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182. https://doi.org/10.1016/j.cell.2020.06.043, 812-827.e19.

Michel, C.J., Mayer, C., Poch, O., Thompson, J.D., 2020. Characterization of accessory genes in coronavirus genomes. Virol. J. 17, 131. https://doi.org/10.1186/s12985-020-01402-1.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Naqvi, A.A.T., Fatima, K., Mohammad, T., Fatima, U., Singh, I.K., Singh, A., Atif, S.M., Hariprasad, G., Hasan, G.M., Hassan, Md.I., 2020. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. Biochim. Biophys. Acta (BBA) - Mol. Basis Dis. 1866, 165878. https://doi.org/10.1016/j.bbadis.2020.165878.

Parvez, Md.S.A., Rahman, M.M., Morshed, Md.N., Rahman, D., Anwar, S., Hosen, M.J., 2021. Genetic analysis of SARS-CoV-2 isolates collected from Bangladesh: insights into the origin, mutational spectrum and possible pathomechanism. Comput. Biol. Chem. 90, 107413. https://doi.org/10.1016/j.compbiolchem.2020.107413.

Planas, D., Bruel, T., Grzelak, L., Guivel-Benhassine, F., Staropoli, I., Porrot, F., Planchais, C., Buchrieser, J., Rajah, M.M., Bishop, E., Albert, M., Donati, F., Prot, M., Behillil, S., Enouf, V., Maquart, M., Smati-Lafarge, M., Varon, E., Schortgen, F., Yahyaoui, L., Gonzalez, M., De Sèze, J., Péré, H., Veyer, D., Sève, A., Simon-Lorière, E., Fafi-Kremer, S., Stefic, K., Mouquet, H., Hocqueloux, L., van der Werf, S., Prazuck, T., Schwartz, O., 2021. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. Nat. Med. 1–8. https://doi.org/10.1038/s41591-021-01318-5.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/ [Internet].

Sagulenko, P., Puller, V., Neher, R.A., 2018. TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol. 4 https://doi.org/10.1093/ve/vex042.

Saha, S., Tanmoy, A.M., Hooda, Y., Tanni, A.A., Goswami, S., Sium, S.M.A., Sajib, M.S.I., Malaker, R., Islam, S., Rahman, H., Anik, A.M., Sarker, N., Islam, M.S., Ghosh, K., Sarkar, P.K., Bipul, M.R.A., Ahmed, S.S., Shahidullah, M., Saha, S.K., 2021. COVID-19 rise in Bangladesh correlates with increasing detection of B.1.351 variant. BMJ Glob. Health 6, e006012. https://doi.org/10.1136/bmjgh-2021-006012.

Salyer, S.J., Maeda, J., Sembuche, S., Kebede, Y., Tshangela, A., Moussif, M., Ihekweazu, C., Mayet, N., Abate, E., Ouma, A.O., Nkengasong, J., 2021. The first and second waves of the COVID-19 pandemic in Africa: a cross-sectional study. Lancet 397, 1265–1275. https://doi.org/10.1016/S0140-6736(21)00632-2.

Scully, E.P., Haverfield, J., Ursin, R.L., Tannenbaum, C., Klein, S.L., 2020. Considering how biological sex impacts immune responses and COVID-19 outcomes. Nat. Rev. Immunol. 20, 442–447. https://doi.org/10.1038/s41577-020-0348-8.

Shishir, T.A., Naser, I.B., Faruque, S.M., 2021. In silico comparative genomics of SARS-CoV-2 to determine the source and diversity of the pathogen in Bangladesh. PLoS One 16, e0245584. https://doi.org/10.1371/journal.pone.0245584.

Shu, Y., McCauley, J., 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull. 22, 30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.

Siam, M.H.B., Hasan, M.M., Raheem, M.E., Khan, H.R., Siddiqee, M.H., Hossain, M.S., 2020. Insights into the first wave of the COVID-19 pandemic in Bangladesh: lessons learned from a high-risk country. medRxiv. https://doi.org/10.1101/2020.08.05.20168674, 2020.08.05.20168674.

Siddiqe, R., Ghosh, A., 2021. Genome-wide in silico identification and characterization of simple sequence repeats in diverse completed SARS-CoV-2 genomes. Gene Rep. 23, 101020. https://doi.org/10.1016/j.genrep.2021.101020.

Takahashi, T., Ellingson, M.K., Wong, P., Israelow, B., Lucas, C., Klein, J., Silva, J., Mao, T., Oh, J.E., Tokuyama, M., Lu, P., Venkataraman, A., Park, A., Liu, F., Meir, A., Sun, J., Wang, E.Y., Casanovas-Massana, A., Wyllie, A.L., Vogels, C.B.F., Earnest, R., Lapidus, S., Ott, I.M., Moore, A.J., Shaw, A., Fournier, J.B., Odio, C.D., Farhadian, S., Cruz, C.D., Grubaugh, N.D., Schulz, W.L., Ring, A.M., Ko, A.I., Omer, S.B., Iwasaki, A., 2020. Sex differences in immune responses that underlie COVID-19 disease outcomes. Nature 588, 315–320. https://doi.org/10.1038/s41586-020-2700-3.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A.J., Engelbrecht, S., Van Zyl, G., Preiser, W., Petruccione, F., Sigal, A., Hardie, D., Marais, G., Hsiao, N., Korsman, S., Davies, M.-A., Tyers, L., Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi, A., Godzik, A., Wibmer, C.K., Sewell, B.T., Lourenço, J., Alcantara, L.C.J., Kosakovsky Pond, S.L., Weaver, S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., de Oliveira, T., 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature 592, 438–443. https://doi.org/10.1038/s41586-021-03402-9.

Van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA. Available from: https://dl.acm.org/doi/book/10.5555/1593511 [Internet].

Volz, E., Hill, V., McCrone, John T., Price, A., Jorgensen, D., O'Toole, Á., Southgate, J., Johnson, Robert, Jackson, B., Nascimento, F.F., Rey, S.M., Nicholls, S.M., Colquhoun, R.M., Filipe, A. da S., Shepherd, J., Pascall, D.J., Shah, R., Jesudason, N., Li, K., Jarrett, R., Pacchiarini, N., Bull, M., Geidelberg, L., Siveroni, I., Koshy, C., Wise, E., Cortes, Nick, Lynch, J., Kidd, S., Mori, M., Fairley, D.J., Curran, T., McKenna, J.P., Adams, H., Fraser, C., Golubchik, T., Bonsall, D., Moore, Catrin, Caddy, S.L., Khokhar, F.A., Wantoch, M., Reynolds, N., Warne, B., Maksimovic, J., Spellman, K., McCluggage, K., John, M., Beer, R., Afifi, S., Morgan, S., Marchbank, A., Price, A., Kitchen, C., Gulliver, H., Merrick, I., Southgate, J., Guest, M., Munn, R., Workman, T., Connor, T.R., Fuller, W., Bresner, C., Snell, L.B., Charalampous, T., Nebbia, G., Batra, R., Edgeworth, J., Robson, S.C., Beckett, A., Loveson, K.F., Aanensen, D.M., Underwood, A.P., Yeats, C.A., Abudahab, K., Taylor, B.E.W., Menegazzo, M., Clark, G., Smith, W., Khakh, M., Fleming, V.M., Lister, M.M., Howson-Wells, H.C., Berry, Louise, Boswell, T., Joseph, A., Willingham, I., Bird, P., Helmer, T., Fallon, K., Holmes, C., Tang, J., Raviprakash, V., Campbell, S., Sheriff, N., Loose, M.W., Holmes, N., Moore, Christopher, Carlile, M., Wright, V., Sang, F., Debebe, J., Coll, F., Signell, A.W., Betancor, G., Wilson, H.D., Feltwell, T., Houldcroft, C.J., Eldirdiri, S., Kenyon, A., Davis, T., Pybus, O., du Plessis, L., Zarebski, A., Raghwani, J., Kraemer, M., Francois, S., Attwood, S., Vasylyeva, T., Torok, M.E., Hamilton, W.L., Goodfellow, I.G., Hall, G., Jahun, A.S., Chaudhry, Y., Hosmillo, M., Pinckert, M.L., Georgana, I., Yakovleva, A., Meredith, L. W., Moses, S., Lowe, H., Ryan, F., Fisher, C.L., Awan, A.R., Boyes, J., Breuer, J., Harris, K.A., Brown, J.R., Shah, D., Atkinson, L., Lee, J.C.D., Alcolea-Medina, A., Moore, N., Cortes, Nicholas, Williams, R., Chapman, M.R., Levett, L.J., Heaney, J., Smith, D.L., Bashton, M., Young, G.R., Allan, J., Loh, J., Randell, P.A., Cox, A., Madona, P., Holmes, A., Bolt, F., Price, J., Mookerjee, S., Rowan, A., Taylor, G.P., Ragonnet-Cronin, M., Nascimento, F.F., Jorgensen, D., Siveroni, I., Johnson, Rob, Boyd, O., Geidelberg, L., Volz, E.M., Brunker, K., Smollett, K.L., Loman, N.J., Quick, J., McMurray, C., Stockton, J., Nicholls, S., Rowe, W., Poplawski, R., Martinez-Nunez, R.T., Mason, J., Robinson, T.I., O'Toole, E., Watts, J., Breen, C., Cowell, A., Ludden, C., Sluga, G., Machin, N.W., Ahmad, S.S.Y., George, R.P., Halstead, F., Sivaprakasam, V., Thomson, E.C., Shepherd, J.G., Asamaphan, P., Niebel, M.O., Li, K.K., Shah, R.N., Jesudason, N.G., Parr, Y.A., Tong, L., Broos, A., Mair, D., Nichols, J., Carmichael, S.N., Nomikou, K., Aranday-Cortes, E., Johnson, N., Starinskij, I., Filipe, A. da S., Robertson, D.L., Orton, R.J., Hughes, J., Vattipally, S., Singer, J.B., Hale, A.D., Macfarlane-Smith, L.R., Harper, K.L., Taha, Y., Payne, B.A.I., Burton-Fanning, S., Waugh, S., Collins, J., Eltringham, G., Templeton, K.E., McHugh, M.P., Dewar, R., Wastenge, E., Dervisevic, S., Stanley, R., Prakash, R., Stuart, C., Elumogo, N., Sethi, D.K., Meader, E.J., Coupland, L.J., Potter, W., Graham, C., Barton, E., Padgett, D., Scott, G., Swindells, E., Greenaway, J., Nelson, A., Yew, W.C., Silva, P.C.R., Andersson, M., Shaw, R., Peto, T., Justice, A., Eyre, D., Crooke, D., Hoosdally, S., Sloan, T.J., Duckworth, N., Walsh, S., Chauhan, A.J., Glaysher, S., Bicknell, K., Wyllie, S., Butcher, E., Elliott, S., Lloyd, A., Impey, R., Levene, N., Monaghan, L., Bradley, D.T., Allara, E., Pearson, C., Muir, P., Vipond, I.B., Hopes, R., Pymont, H.M., Hutchings, S., Curran, M.D., Parmar, S., Lackenby, A., Mbisa, T., Platt, S., Miah, S., Bibby, D., Manso, C., Hubb, J., Chand, M., Dabrera, G., Ramsay, M., Bradshaw, D., Thornton, A., Myers, R., Schaefer, U., Groves, N., Gallagher, E., Lee, D., Williams, D., Ellaby, N., Harrison, I., Hartman, H., Manesis, N., Patel, V., Bishop, C., Chalker, V., Osman, H., Bosworth, A., Robinson, E., Holden, M.T.G., Shaaban, S., Birchley, A., Adams, A., Davies, A., Gaskin, A., Plimmer, A., Gatica-Wilcox, B., McKerr, C., Moore, Catherine, Williams, C., Heyburn, D., Lacy, E.D., Hilvers, E., Downing, F., Shankar, G., Jones, H., Asad, H., Coombes, J., Watkins, J., Evans, J.M., Fina, L., Gifford, L., Gilbert, L., Graham, L., Perry, M., Morgan, M., Bull, M., Cronin, M., Pacchiarini, N., Craine, N., Jones, R., Howe, R., Corden, S., Rey, S., Kumziene-Summerhayes, S., Taylor, S., Cottrell, S., Jones, S., Edwards, S., O'Grady, J., Page, A.J., Wain, J., Webber, M.A., Mather, A.E., Baker, D.J., Rudder, S., Yasir, M., Thomson, N.M., Aydin, A., Tedim, A.P., Kay, G.L., Trotter, A.J., Gilroy, R.A.J., Alikhan, N.-F., Martins, L. de O., Le-Viet, T., Meadows, L., Kolyva, A., Diaz, M., Bell, A., Gutierrez, A.V., Charles, I.G., Adriaenssens, E.M., Kingsley, R.A., Casey, A., Simpson, D.A., Molnar, Z., Thompson, T., Acheson, E., Masoli, J.A.H., Knight, B.A., Hattersley, A., Ellard, S., Auckland, C., Mahungu, T.W., Irish-Tavares, D., Haque, T., Bourgeois, Y., Scarlett, G.P., Partridge, D.G., Raza, M., Evans, C., Johnson, K., Liggett, S., Baker, P., Essex, S., Lyons, R.A., Caller, L.G., Castellano, S., Williams, R.J., Kristiansen, M., Roy, S., Williams, C.A., Dyal, P.L., Tutill, H.J., Panchbhaya, Y.N., Forrest, L.M., Niola, P., Findlay, J., Brooks, T.T., Gavriil, A., Mestek-Boukhibar, L., Weeks, S., Pandey, S., Berry, Lisa, Jones, K., Richter, A., Beggs, A., Smith, C.P., Bucca, G., Hesketh, A.R., Harrison, E.M., Peacock, S.J., Palmer, Sophie, Churcher, C. M., Bellis, K.L., Girgis, S.T., Naydenova, P., Blane, B., Sridhar, S., Ruis, C., Forrest, S., Cormie, C., Gill, H.K., Dias, J., Higginson, E.E., Maes, M., Young, J., Kermack, L.M., Hadjirin, N.F., Aggarwal, D., Griffith, L., Swingler, T., Davidson, R.K., Rambaut, A., Williams, T., Balcazar, C.E., Gallagher, M.D., O'Toole, Á., Rooke, S., Jackson, B., Colquhoun, R., Ashworth, J., Hill, V., McCrone, J.T., Scher, E., Yu, X., Williamson, K. A., Stanton, T.D., Michell, S.L., Bewshea, C.M., Temperton, B., Michelsen, M.L., Warwick-Dugdale, J., Manley, R., Farbos, A., Harrison, J.W., Sambles, C.M., Studholme, D.J., Jeffries, A.R., Darby, A.C., Hiscox, J.A., Paterson, S., Iturriza-Gomara, M., Jackson, K.A., Lucaci, A.O., Vamos, E.E., Hughes, M., Rainbow, L., Eccles, R., Nelson, C., Whitehead, M., Turtle, L., Haldenby, S.T., Gregory, R., Gemmell, M., Kwiatkowski, D., de Silva, T.I., Smith, N., Angyal, A., Lindsey, B.B., Groves, D.C., Green, L.R., Wang, D., Freeman, T.M., Parker, M.D., Keeley, A.J., Parsons, P.J., Tucker, R.M., Brown, R., Wyles, M., Constantinidou, C., Unnikrishnan, M., Ott, S., Cheng, J.K.J., Bridgewater, H.E., Frost, L.R., Taylor-Joyce, G., Stark, R., Baxter, L., Alam, M.T., Brown, P.E., McClure, P.C., Chappell, J. G., Tsoleridis, T., Ball, J., Gramatopoulos, D., Buck, D., Todd, J.A., Green, A., Trebes, A., MacIntyre-Cockett, G., de Cesare, M., Langford, C., Alderton, A., Amato, R., Goncalves, S., Jackson, D.K., Johnston, I., Sillitoe, J., Palmer, Steve, Lawniczak, M., Berriman, M., Danesh, J., Livett, R., Shirley, L., Farr, B., Quail, M., Thurston, S., Park, N., Betteridge, E., Weldon, D., Goodwin, S., Nelson, R., Beaver, C., Letchford, L., Jackson, D.A., Foulser, L., McMinn, L., Prestwood, L., Kay, S., Kane, L., Dorman, M.J., Martincorena, I., Puethe, C., Keatley, J.-P., Tonkin-Hill, G., Smith, C., Jamrozy, D., Beale, M.A., Patel, M., Ariani, C., Spencer-Chapman, M., Drury, E., Lo, S., Rajatileka, S., Scott, C., James, K., Buddenborg, S.K., Berger, D.J., Patel, G., Garcia-Casado, M.V., Dibling, T., McGuigan, S., Rogers, H.A., Hunter, A.D., Souster, E., Neaverson, A.S., Goodfellow, I., Loman, N.J., Pybus, O.G., Robertson, D. L., Thomson, E.C., Rambaut, A., Connor, T.R., 2021a. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell 184. https://doi.org/10.1016/j.cell.2020.11.020, 64-75.e11.

Volz, E., Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C.V., Boyd, O., Loman, N.J., McCrone, J.T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D.K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D.P., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., Ferguson, N.M., 2021b. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature 1–6. https://doi.org/10.1038/s41586-021-03470-x.

Wibmer, C.K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Oosthuysen, B., Lambson, B.E., de Oliveira, T., Vermeulen, M., van der Berg, K., Rossouw, T., Boswell, M., Ueckermann, V., Meiring, S., von Gottberg, A., Cohen, C., Morris, L., Bhiman, J.N., Moore, P.L., 2021. SARS-CoV-2 501Y.V2 escapes neutralization by south African COVID-19 donor plasma. Nat. Med. 27, 622–625. https://doi.org/10.1038/s41591-021-01285-x.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis, 2nd ed. 2016. ed. Springer, Cham. https://doi.org/10.1007/978-3-319-24277-4. Use R! Springer International Publishing: Imprint.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. J. Open Source Softw. 4, 1686. Doi:10.21105/joss.01686.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E.C., Zhang, Y.-Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269. https://doi.org/10.1038/s41586-020-2008-3.

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T.P., Wang, Y., Baum, A., Diehl, W.E., Dauphin, A., Carbone, C., Veinotte, K., Egri, S.B., Schaffner, S. F., Lemieux, J.E., Munro, J.B., Rafique, A., Barve, A., Sabeti, P.C., Kyratsous, C.A., Dudkina, N.V., Shen, K., Luban, J., 2020. Structural and functional analysis of the

D614G SARS-CoV-2 spike protein variant. Cell 183. https://doi.org/10.1016/j.cell.2020.09.032, 739-751.e8.

Zhou, D., Dejnirattisai, W., Supasa, P., Liu, C., Mentzer, A.J., Ginn, H.M., Zhao, Y., Duyvesteyn, H.M.E., Tuekprakhon, A., Nutalai, R., Wang, B., Paesen, G.C., Lopez-Camacho, C., Slon-Campos, J., Hallis, B., Coombes, N., Bewley, K., Charlton, S., Walter, T.S., Skelly, D., Lumley, S.F., Dold, C., Levin, R., Dong, T., Pollard, A.J., Knight, J.C., Crook, D., Lambe, T., Clutterbuck, E., Bibi, S., Flaxman, A., Bittaye, M., Belij-Rammerstorfer, S., Gilbert, S., James, W., Carroll, M.W., Klenerman, P., Barnes, E., Dunachie, S.J., Fry, E.E., Mongkolsapaya, J., Ren, J., Stuart, D.I., Screaton, G.R., 2021. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. Cell 184. https://doi.org/10.1016/j.cell.2021.02.037, 2348–2361.e6.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., Shi, Z.-L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273. https://doi.org/10.1038/s41586-020-2012-7.