VIRUS EVOLUTION

Virus Evolution, 2022, **8(1)**, 1–9
DOI: https://doi.org/10.1093/ve/veac020
Advance access publication date: 17 March 2022
Research Article

Evolutionary history and introduction of SARS-CoV-2 Alpha VOC/B.1.1.7 in Pakistan through international travelers

Asghar Nasir,[1] Ali Raza Bukhari,[1,†] Nídia S. Trovão,[2,†,‡] Peter M. Thielen,[3,§] Akbar Kanji,[1] Syed Faisal Mahmood,[4] Najia Karim Ghanchi,[1] Zeeshan Ansar,[1] Brian Merritt,[3] Thomas Mehoke,[3,¶] Safina Abdul Razzak,[1,**] Muhammed Asif Syed,[5] Suhail Raza Shaikh,[5] Mansoor Wassan,[5] Uzma Bashir Aamir,[6] Guy Baele,[7,††] Zeba Rasmussen,[2,§§] David Spiro,[2] Rumina Hasan,[1] and Zahra Hasan[1,*,¶¶]

[1]Department of Pathology and Laboratory Medicine, The Aga Khan University, Stadium Road, Karachi, Pakistan, [2]Fogarty International Center, U.S. National Institutes of Health, 16 Center Drive, Bethesda, MD 20892, USA, [3]Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA, [4]Department of Medicine, The Aga Khan University, Karachi, Stadium Road, Pakistan, [5]Department of Health, Government of Sindh, Sindh Secretariat, Kamall Atta Turk Road, Karachi, Pakistan, [6]World Health Organization country office, Park Road, Chak Shahzad, Islamabad, Pakistan and [7]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium
[†]Authors contributed equally to this work.
[‡]https://orcid.org/0000-0002-2106-1166
[§]https://orcid.org/0000-0003-1807-2785
[¶]https://orcid.org/0000-0001-6607-8925
[**]https://orcid.org/0000-0001-8073-0684
[††]https://orcid.org/0000-0002-1915-7732
[§§]https://orcid.org/0000-0003-1798-781X
[¶¶]https://orcid.org/0000-0001-7580-372X
*Corresponding author: E-mail: zahra.hasan@aku.edu

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants continue to emerge, and their identification is important for the public health response to coronavirus disease 2019 (COVID-19). Genomic sequencing provides robust information but may not always be accessible, and therefore, mutation-based polymerase chain reaction (PCR) approaches can be used for rapid identification of known variants. International travelers arriving in Karachi between December 2020 and February 2021 were tested for SARS-CoV-2 by PCR. A subset of positive samples was tested for S-gene target failure (SGTF) on TaqPathTM COVID-19 (Thermo Fisher Scientific) and for mutations using the GSD NovaType SARS-CoV-2 (Eurofins Technologies) assays. Sequencing was conducted on the MinION platform (Oxford Nanopore Technologies). Bayesian phylogeographic inference was performed integrating the patients' travel history information. Of the thirty-five COVID-19 cases screened, thirteen had isolates with SGTF. The travelers transmitted infection to sixty-eight contact cases. The B.1.1.7 lineage was confirmed through sequencing and PCR. The phylogenetic analysis of sequence data available for six cases included four B.1.1.7 strains and one B.1.36 and B.1.1.212 lineage isolate. Phylogeographic modeling estimated at least three independent B.1.1.7 introductions into Karachi, Pakistan, originating from the UK. B.1.1.212 and B.1.36 were inferred to be introduced either from the UK or the travelers' layover location. We report the introduction of SARS-CoV-2 B.1.1.7 and other lineages in Pakistan by international travelers arriving via different flight routes. This highlights SARS-CoV-2 transmission through travel, importance of testing, and quarantine post-travel to prevent transmission of new strains, as well as recording detailed patients' metadata. Such results help inform policies on restricting travel from destinations where new highly transmissible variants have emerged.

Key words: B.1.1.7 variant; Pakistan; Bayesian inference; Markov chain Monte Carlo; phylogenetics; phylodynamics.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in Wuhan in December 2019 (Huang et al. 2020; Wu et al. 2020). As of 18 September 2021, 228 million infections have been reported globally, with 4.7 million deaths (JHU 2021). The first wave of COVID-19 in Pakistan occurred between March and July 2020 (JHU 2021) and the second from October 2020 until January 2021. The third wave occurred from March to May 2021(Imran et al. 2021) and soon after in July a fourth wave started, which is still ongoing (Pakistan 2021). An estimated 1.5 million cases have been reported in Pakistan with a death toll of 30,000 (https://coronavirus.jhu.edu/region/pakistan, last accessed 9 March 2022).

New variants have continued to emerge through the coronavirus disease 2019 (COVID-19) pandemic. There is limited information regarding the genomic epidemiology of SARS-CoV-2

strains in Pakistan. G clade strains were identified in March 2020, and their predominance varied through the year (Ghanchi et al. 2021; Umair et al. 2021). The B.1.1.7 lineage was first identified in Pakistan in January 2021 (Tamim et al. 2021).

SARS-CoV-2 strains with mutations in the spike (S) glycoprotein region D614G, identified as G clade strains (GISAID) or B.1 as per Pango lineage, were first associated with greater transmissibility due to enhanced viral replication in human lung epithelial cells by increasing the infectivity and stability of virions (Plante et al. 2021). The B.1.1.7 lineage is defined by twenty-three mutations that include nonsynonymous mutations or deletions in the *orf*1ab gene (T1001I, A1 708D, 12230T, 3675–3677 deletion), S gene (69–70 deletion, Y144 deletion, N501Y, A570D, P681H, T7161I, S982A, D111BH), *orf*8 gene (Q27 STOP, R52I, Y73C0), nucleocapsid (N) gene (D3L, S235F0), and synonymous mutations in the *orf*1ab and membrane (M) genes (Frampton et al. 2021).

The B.1.1.7 variant was subsequently correlated with a significant increase in the rate of COVID-19 infections in the UK (Challen et al. 2021). The increased transmissibility, morbidity, and mortality associated with the B.1.1.7 lineage makes a case for monitoring its dynamics in the population. This VOC includes a deletion in the S gene that results in S gene dropout or S-gene target failure (SGTF) using the TaqPath COVID-19 CE-IVD RT-PCR (Thermo Fisher Scientific) (Bal et al. 2021). In the UK, Public Health England (PHE) screened isolates using this assay followed by genome sequencing-based confirmation to show 96 per cent of samples with SGFT in January 2021 to be B.1.1.7 lineage isolates (PHE 2021).

Given the rising concern of the spread of the B.1.1.7 variant in the UK, the Government of Pakistan put temporary travel restrictions in place for travelers arriving from the UK in the last week of December 2020. After this time, international passengers coming into Pakistan were required to have a negative SARS-CoV-2 polymerase chain reaction (PCR) test within 72 hours prior to their travel. A 7-day home quarantine guidance was put in place and PCR testing was conducted for household contacts of COVID-19 cases. On arrival, mandatory PCR testing was conducted for all inbound passengers from the end of December 2020 until February 2021.

In this study, we report the results of COVID-19 screening of travelers arriving in Pakistan from the UK during the period from December 2020 through February 2021. A phylogeographic analysis that incorporates the individual travel history of sampled patients confirmed repeated introductions and transmission of the B.1.1.7 lineage in Karachi, Pakistan.

2. Materials and Methods

2.1 Ethical approval

This study was approved by the Ethical Review Committee, The Aga Khan University, Pakistan (ERC#2020-4732-10292).

2.2 SARS-CoV-2 diagnostics

The Aga Khan University Hospital (AKUH) Clinical Laboratories are accredited by the College of American Pathologists, USA. Routine testing for SARS-CoV-2 by PCR testing was performed on the cobas® SARS-CoV-2 assay (Roche Diagnostics) in accordance with the manufacturer's recommendations. AKUH worked with the Department of Health, Government of Sindh, through the COVID-19 pandemic to provide rapid, clinical diagnostic testing for suspected cases. Results of specimens received for SARS-CoV-2 PCR testing were reported to the health authorities on a daily basis.

2.3 Selection of study specimens

As part of public health surveillance, respiratory specimens from passengers arriving at Karachi airport on multiple flights from the UK between the period 24 December 2020 and 15 February 2021 were submitted to AKUH for SARS-CoV-2 testing by the Health Department, Government of Sindh. Here we selected a subset of thirty-five specimens with a crossing threshold (Ct) value less than thirty on the cobas® SARS-CoV-2 assay.

2.4 SARS-CoV-2 testing using the TaqPath COVID-19 assay

Selected SARS-CoV-2 specimens were screened for SGTF on the TaqPath COVID-19 (Thermofisher, USA) assay as per the manufacturer's recommendations. Briefly, RNA was extracted from respiratory specimens using the Qiagen RNA MiniPrep kit (Qiagen, USA). Amplification of N, *orf*1ab, and S-genes was performed and samples with SGTF were identified. Representative profiles of samples with SGTF and non-SGTF are shown in Supplementary Figure S1.

2.5 Targeted PCR for B.1.1.7 lineage-defining mutations

Specimens were tested with the GSD NovaType SARS-CoV-2 assay (Eurofins Technologies); samples were screened for N501Y/A570D mutations. The assay detects the two mutations at nucleotide positions A23063T and C23271A, respectively. Lineage B.1.1.7 carries both mutations, whereas lineages B.1.351 and P.1 carry only N501Y. Detection of both mutations by using different fluorophores allows discrimination between wild-type and SARS-CoV-2 variants.

2.6 SARS-CoV-2 sequencing and variant analysis

2.6.1 Genome sequencing

A total of eight samples were processed for sequencing using the ARTIC network protocol, V.3 (Quick 2019) on the Oxford Nanopore Technologies (ONT) MinION Mk1B device. Briefly, viral RNA was reverse transcribed to complementary (c)DNA using SuperScript IV Vilo Master Mix (Thermo Inc., USA). Tiled Primers for the SARS-CoV-2 amplifications designed by the ARTIC network were used to amplify the viral genome using primer pools A and B. For the amplification from each pool of primers, we used 3 μl of reverse-transcribed viral RNA, 12.5 μl of Q5 DNA master mix (NEB, USA), and 1 μl (10 μM) of primer for a reaction mix of 25 μl.

The amplification program was run at 98°C for 30 seconds, followed by 30 cycles at 95°C for 15 seconds and 65 °C for 5 minutes. The reaction was then held at 4°C. PCR products from pools A and B were combined and then purified using magnetic beads. About 50 ng of purified DNA proceeded to end repair using 1.75 μl of Ultra II End Prep Reaction Buffer and 0.75 μl of Ultra II End Prep Enzyme Mix. End repair mix was incubated at 20°C for 10 minutes followed by heat inactivation at 65°C for 10 seconds. Samples were then held at 4°C until further processing. The barcoded adapters were ligated using the native barcode expansion kit (EXP-NBD-114) from ONT following the manufacturer's instructions. After adapter ligation, samples were purified and equal concentrations of samples were pooled together. The pooled samples were purified and motor protein ligation was performed as per ONT standard protocol. About 10 ng of DNA was loaded onto the MinION flow cell (FLO-MIN106D). Basecalling was performed using high-accuracy mode (Guppy v4.3.4) and FASTQ files from the FASTQ pass folder were used for further analysis.

2.7 Variant analysis

Data were aligned with BWA-MEM v.0.7.17r1188 to the Wuhan-Hu-1 reference (Genbank MN908947.3). Sequence Alignment Map (SAM) and Binary Alignment Map (BAM) Files were generated using SAMtools v.1.10.2-3 and variants were called with BCFtools (Binary and Variant Call Format tools) v.1.10.2 with mpileup. Fastall (FASTA) generation was done using Medeka and Nanopolish incorporated in a python script that took the output from both tools and made a consensus FASTA based on the generated results. The sequences were deposited on GISAID with accession numbers as follows: S5 (EPI_ISL_2608415), S6 (EPI_ISL_2608416), S7 (EPI_ISL_2608417), S10 (EPI_ISL_2608418), S11 (EPI_ISL_2608419), S12 (EPI_ISL_2696298), and S20 (EPI_ISL_2068420) (Supplementary Table S1). For further analysis at the low-coverage regions, we imported the sorted BAM files into the Integrative Genomics Viewer v. 2.8.13 and performed manual inspection of reads at particular nucleotide positions.

2.8 Evolutionary analysis

2.8.1 Phylogenetic classification through lineages

We used the phylogenetic assignment of named global outbreak lineages (PANGOLIN; Rambaut et al. 2020; Toole et al. 2021) to evaluate the genetic diversity of the eight genetic sequences generated along with global lineages.

2.9 Selecting a genomic background dataset

For phylogenetic analyses, full-length viral genome sequences belonging to Pango lineages identified above (B.1.1.7, B.1.1.212, and B.1.36) were downloaded from GISAID (Elbe and Buckland-Merrett 2017) on 6 February 2021. Multiple sequence alignment was performed using MAFFT v7.458 44 using parameters –reorder –any symbol –nomemsave –adjust direction –add fragments and used Wuhan-Hu-1 (GenBank accession number: MN908947.3) sequence as a reference (Katoh, Rozewicki, and Yamada 2019). Sequences with fewer than 75 per cent unambiguous bases were excluded, as were duplicate sequences defined as having identical nucleotide composition and having been collected on the same date and in the same country, since these do not contribute to the overall representation of viral diversity. The resulting dataset was trimmed at the 5′ and 3′ ends resulting in a multiple sequence alignment of 29,782 nucleotides.

For computational efficiency, we subsampled the lineage-specific datasets homogeneously through time and space (for the B.1.1.212 dataset, we selected 17 sequences/country/year; for the B.1.36 dataset, we selected 20 sequences/country/year; for the B.1.1.7 dataset, we selected 45 sequences/region/year), resulting in datasets that represent the globally circulating diversity. We preferentially selected longer sequences with the fewest number of gaps in the 5′ and 3′ ends and those that had complete dates and the fewest number of ambiguous bases. These datasets were subjected to multiple iterations of maximum-likelihood phylogenetic reconstruction using IQ-TREE v2.1.3 (Nguyen et al. 2015) with parameters -m GTR + G -nt 50 and exclusion of outlier sequences whose genetic divergence and sampling date were incongruent using TempEst (Rambaut et al. 2016). Isolates S12 and S21 that were generated in this study were excluded from subsequent phylodynamic analysis, as they behaved as outliers in the root-to-tip regression (Supplementary Figure SX). The final datasets had 475 (B.1.1.7) sequences, 38 (B.1.1.212) sequences, and 550 (B.1.36) sequences, including those sequences that were generated during this study (Supplementary Table S2).

2.10 Phylodynamic reconstruction

The evolutionary history and spatiotemporal dynamics were inferred for the final datasets using a Bayesian phylogeographic approach using Markov chain Monte Carlo (MCMC) available via the BEAST v1.10.5 software package (Suchard et al. 2018) and using the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD, USA (http://biowulf.nih.gov).

Our interest lies in estimating the viral evolutionary history and spatial diffusion process of the introduction events of the B.1.1.7, B.1.1.212, and B.1.36 lineages. Travel history data are of particular importance when analyzing low diversity data, such as that for SARS-CoV-2, using Bayesian joint inference of sequence and location traits because sharing the same location state can contribute to the phylogenetic clustering of taxa (Lemey et al. 2020). For each of the lineage-specific datasets, we performed a joint genealogical and phylogeographic inference of time-measured trees using MCMC as implemented in the BEAST v1.10.5 software package (Suchard et al. 2018). We assumed a general-time reversible (GTR) model of nucleotide substitution with gamma-distributed rate variation among sites (Tavaré 1985; Yang 1994; Edwards et al. 2011). We used an uncorrelated log-normal relaxed molecular clock to account for evolutionary rate variation on the branches of the phylogeny (Drummond et al. 2006) and specified an exponential growth coalescent prior in our analyses in which effective population size and rate of exponential growth are estimated. We used the default priors in BEAUti, with the exception of the exponential population size (log-normal distribution; mu = 1.0; sigma = 1.5) and exponential growth size (Laplace distribution; mean = 0.0; standard deviation = 100.0). For sequences with only the year of viral collection available, the lack of tip date precision was accommodated by sampling uniformly across a 1-year window from 1 January to 30 December 2020.

To integrate the travel history information obtained from (returning) travelers, including layover information, we followed Lemey et al. (2020) and augmented the phylogeny with ancestral nodes that are associated with a location state (but not with a known sequence) and enforced the ancestral location at a point in the past of a lineage. We specified normal prior distributions on the travel times informed by an estimate of time of infection and truncated to be positive (back-in-time) relative to sampling date. Specifically, we used a mean of 10 days before sampling based on a mean incubation time of 5 days (Lauer et al. 2020) and a constant ascertainment period of 5 days between symptom onset and testing (Ferguson et al. 2020) and a standard deviation of 3 days to incorporate the uncertainty on the incubation time. The location traits associated with taxa and with the ancestral nodes were modeled using a bidirectional asymmetric discrete diffusion process (Lemey et al. 2009). The MCMC analysis was run separately at least five times for each of the datasets and for at least 200 million iterations with subsampling every 20,000 iterations using the BEAGLE (Ayres et al. 2012) library to improve computational performance. All parameters reached convergence and showed adequate statistical mixing effect sample size (ESS > 200), as assessed visually using Tracer v.1.7.1 (Rambaut et al. 2018) with statistical uncertainty reflected in values of the 95 per cent highest posterior density. At least 10 per cent of the chain was removed as burn-in. Maximum clade credibility (MCC) trees were summarized using TreeAnnotator v1.10.5 and visualized in FigTree v1.4.4.

2.11 Statistical analysis

Statistical analyses were performed using Graphpad Prism 5.0. The data for Ct values are expressed as medians with their ranges and standard errors. The significance of the differences was calculated using a nonparametric Mann–Whitney test. $P < 0.05$ was considered to indicate statistical significance.

3. Results

3.1 SGTF screening cases among UK travelers and their SARS-CoV-2 viral load analysis

We analyzed specimens from thirty-five COVID-19-positive travelers. The individuals had an average age of 36 years (range 4–70 years). Six individuals were aged under 18 years, seventeen were aged 19–44 years, and twelve were >45 years of age. Of the thirty-five samples, thirteen were found to have the S-gene deletion 21765–21770 (HV 69–70 deletion) resulting in SGTF on the TaqPath COVID-19 assay. The remaining twenty-two cases were non-SGTF (Supplementary Table S3). There was no statistically significant difference between Ct values of SARS-CoV-2 strains diagnosed with SGTF and non-SFTF (Fig. 1).

3.2 Transmission of SARS-CoV-2 from travelers to local contacts

All COVID-19 cases among travelers ($n = 35$) were reported directly to the Health Department, who subsequently screened their household contacts for SARS-CoV-2 infection. SARS-CoV-2 PCR test results of the household contacts were available for thirty-four such index cases (Supplementary Table S4). In total, they had 206 household contacts of whom 68 (33 per cent) were reported to be positive for SARS-CoV-2. The median number of additional COVID-19-positive cases in each household was 2, ranging between zero and six individuals. Of note, traveler S11 had an additional six positive household contacts. As thirty-four travelers caused sixty-eight additional COVID-19 infections.

We further determined whether there was any difference in the number of SARS-CoV-2 infections caused by individuals infected with strains that did or did not have S gene target failure (SGTF) or B.1.1.7 variants. There were thirteen index cases with SGTF and twenty-one with non-SGTF. They caused twenty-four and forty-four COVID-19 transmissions, respectively. There was no significant difference found between the transmission rate of SARS-CoV-2 between SGTF and non-SGTF index cases.

3.3 Sequencing of strains

Eight SARS-CoV-2 genomes, consisting of six SGTF and two non-SGTF cases, were successfully sequenced. The six isolates with SGTF were revealed to be B.1.1.7 strains based on their signature lineage-defining mutations. The two non-SGTF isolates were found to belong to B.1.36 and B.1.1.212 lineages. Details of the lineage-defining polymorphisms for the isolates can be found in Supplementary Figure S2. The analysis of the sequences reveals the presence/absence of key differences between the isolates particular to the regions *orf*1ab-gene deletions at positions 11288–11296 (Supplementary Figure S3A); S-gene deletions at positions 21765–21770 (Supplementary Figure S3B) and 21991–21993 (Supplementary Figure S3C) and N-gene GAT CTA conversion at positions 28280–28282 (Supplementary Fig S3D).

3.4 Targeted PCR testing of B.1.1.7 lineage variants (N501Y/A570D)

Nine samples with SGTF, which could not be sequenced, were further investigated by analyzing N501Y/A570D mutations using the NovaType SARS-CoV-2 assay. All of the nine samples tested were positive for N501Y and A570D mutations at nucleotide positions A23063T and C23271A, respectively (Supplementary Figure S2), thereby confirming that they were B.1.1.7 lineage strains.

3.5 Evolutionary and spatiotemporal patterns of SARS-CoV-2 in Pakistan

In order to assess the origins and measure the number of SARS-CoV-2 introductions into Karachi, Pakistan, during the

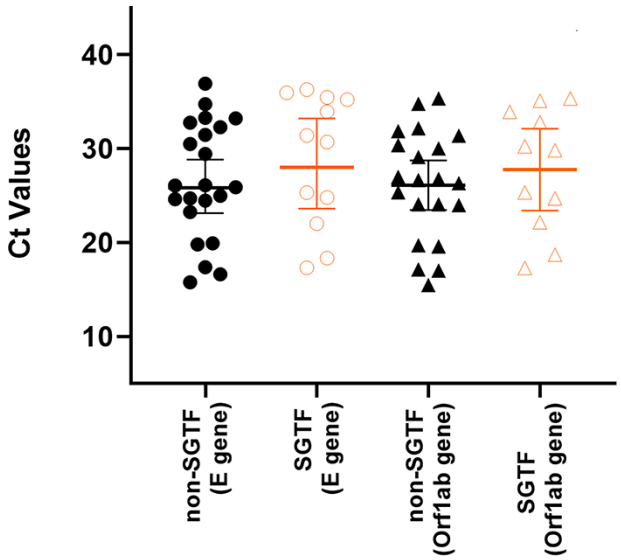

Figure 1. Comparison of viral loads of SGTF and non-SGTF samples. Nasopharyngeal swab samples from travelers arriving from the UK were screened on the cobas® SARS-CoV-2 PCR assay (Roche). Subsequently, all SARS-CoV-2 positive samples were screened for SGTF on the TaqPath COVID-19. For comparison of viral loads on SGTF and non-SGTF sample, the Ct values of cobas Roche SARS-CoV-2 PCR assay (Target 1—ORF1ab gene) and Target 2 (E-gene) were analyzed. There was no statistically significant difference found between Ct values of SGTF and non-SGTF strains for either E or Orf1ab gene amplification.

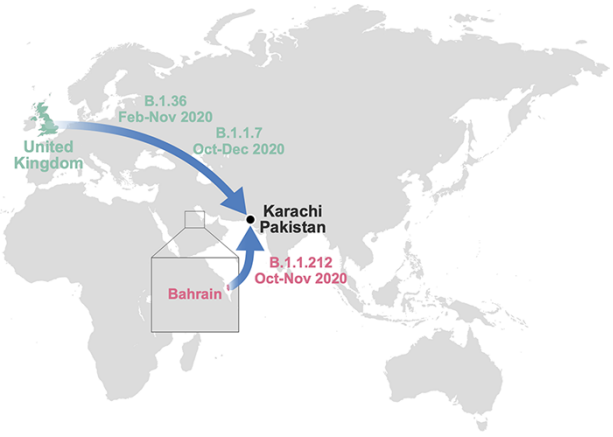

Figure 2. Spatial dispersal of B.1.1.221, B.1.36, and B.1.1.7 lineages into Karachi, Pakistan. The spatial dispersal of SARS-CoV-2 viruses from Bahrain (magenta) and the UK (green) into Karachi, Pakistan, was inferred from the respective MCC trees (Figs 4–6). Arrows project the MCC trees' branches that lead to the seeding events of the B.1.1.212, B.1.36, and B1.1.7 lineages in Pakistan.

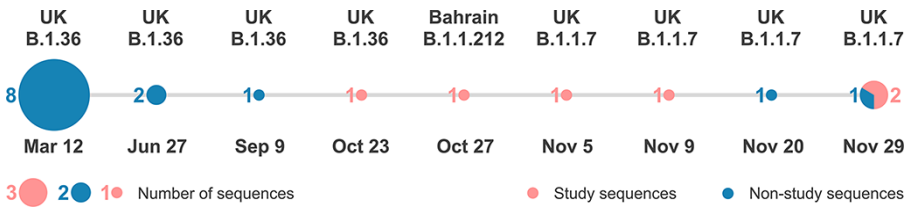

Figure 3. Timeline of introductions of B.1.1.221, B.1.36, and B.1.1.7 lineages into Karachi, Pakistan. Circles represent lineage-specific viral introductions inferred as Pakistani-specific clades in the respective MCC trees (Figs 4–6). The shade depicts if a clade was composed of sequences generated in this study (lighter colour) or previously available (darker colour). The number of sequences in clade is depicted by the number and size of the circles.

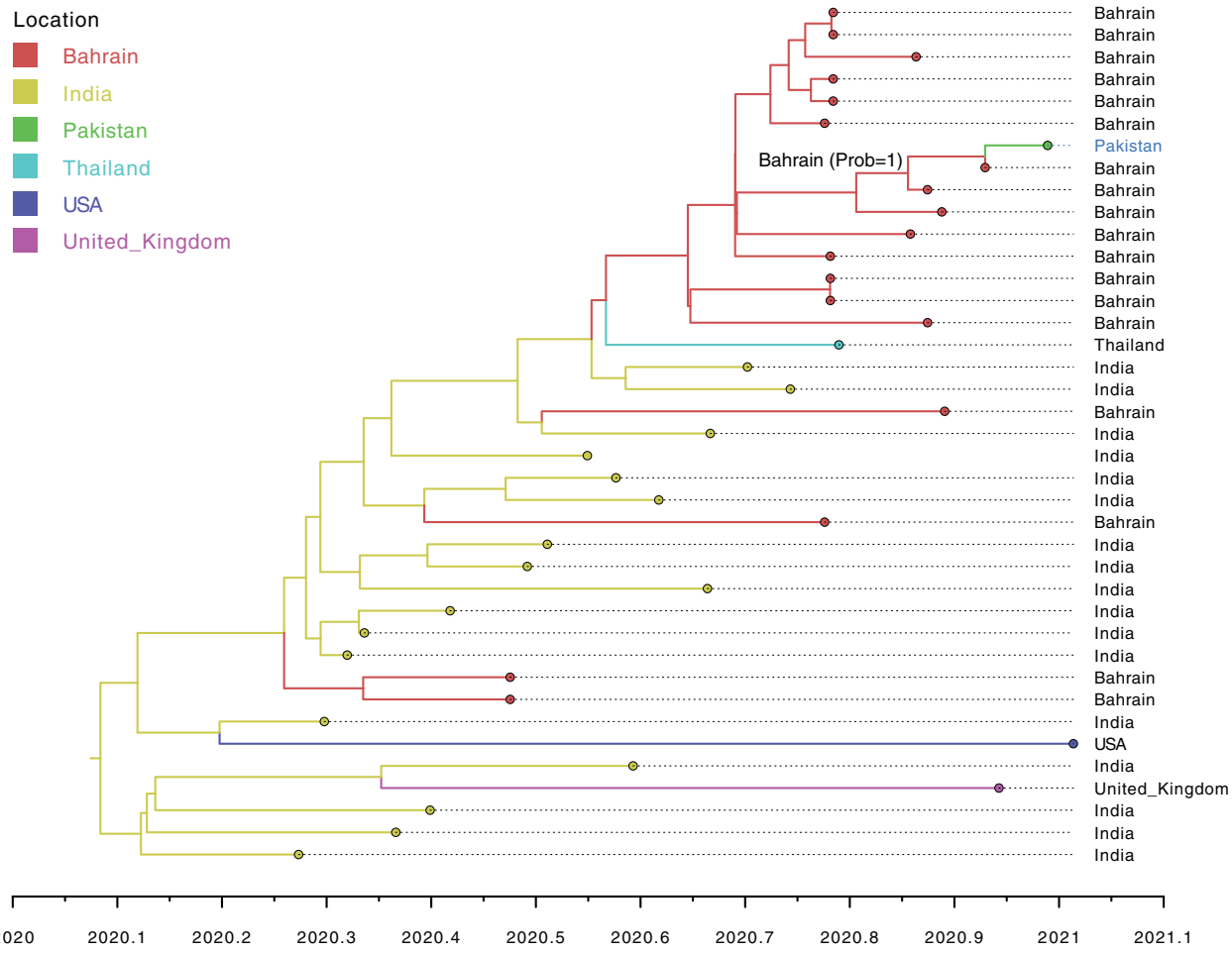

Figure 4. MCC tree inferred for the whole genomes of lineage B.1.1.212. The shade of the branches and tips indicates the inferred location state at the nodes. The ancestral node of run3/s1 (S10) sequence is annotated with the inferred location and probability.

period of December 2020 through February 2021, we performed Bayesian phylodynamic analyses integrating rich travel history data, including layover information, for all isolates included in these analyses (Figs 2 and 3). Overall, we detected nine clades (one from lineage B.1.1.212 (Fig. 4), four from lineage B.1.36 (Figs 5 and 4) and four from lineage B.1.1.7 (Fig. 6)), representing SARS-CoV-2 introductions into Pakistan.

We integrated information regarding the travel history of sequence run3/s1 (S10) to the UK with a layover in Bahrain on their way back to Pakistan. The genetic diversity of this sequence belonging to the B.1.1.212 lineage traces back to a most recent common ancestor that existed between 6 October and 13 November 2020, in Bahrain (location probability = 1) (Figs 2–4). Sequence run3/s5 (S5), belonging to lineage B.1.36, had a record of the travel history to the UK and layover in Qatar. We estimated that the viral introduction that resulted in this infection occurred in the UK (location probability = 0.83), likely between 4 September and 22 November 2020. Other Pakistani sequences belonging to the B.1.36 lineage resulted from three additional introductions are inferred to have an origin in the UK (with location probabilities of 0.39, 0.52, and 0.85) during the period between 14 February and 30 September 2020. One of these introductions was inferred to have generated a transmission chain of at least eight cases (Figs 2, 3, and 5).

Most of the sequences generated in this study belong to Pango lineage B.1.1.7. Sequence run3/s2 (S11) had a record of the travel history to the UK and layover in the UAE. We estimated that the viral introduction that resulted in this infection occurred in the UK (location probability = 0.73), likely between 20 October and

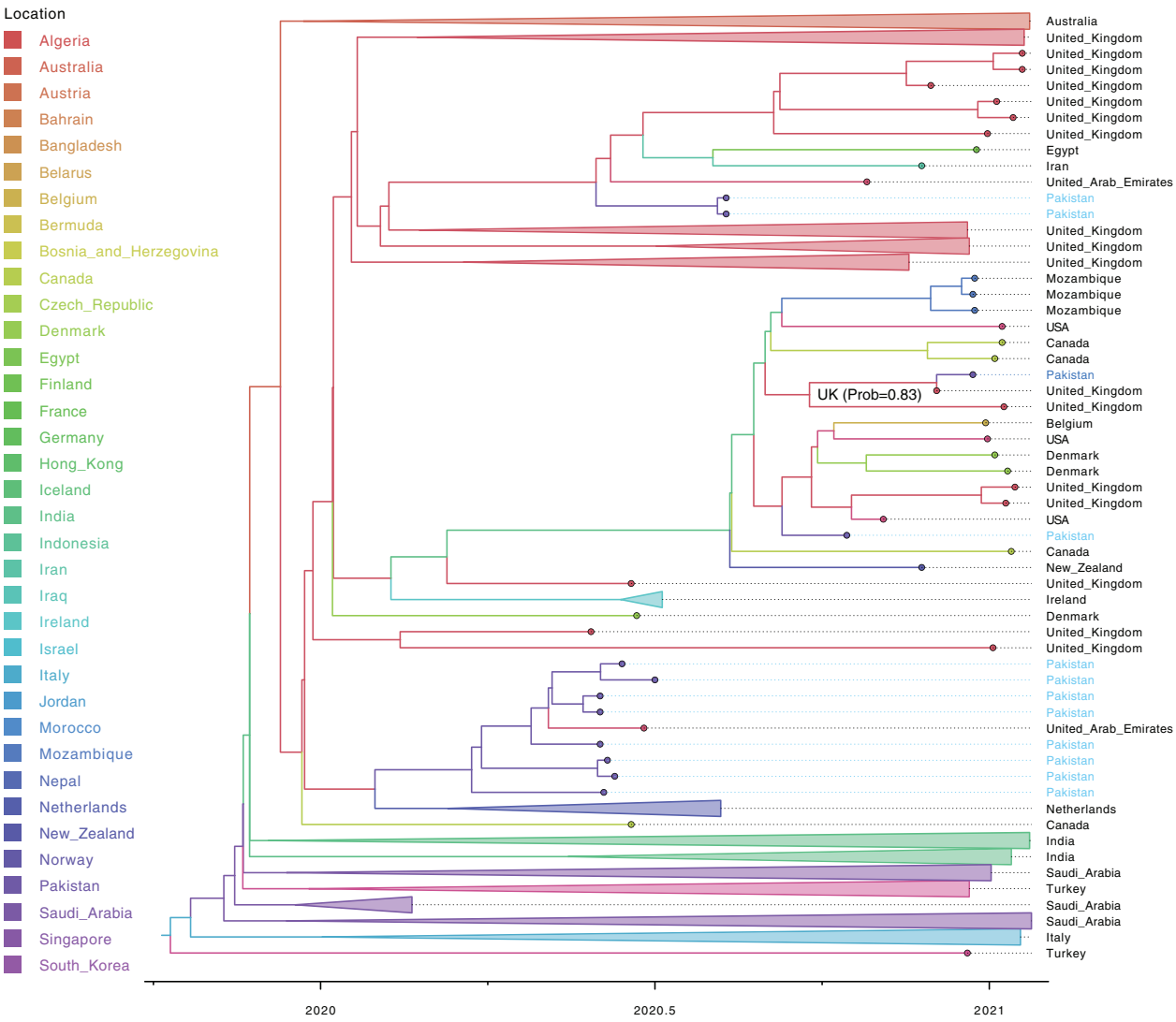

Figure 5. MCC tree inferred for the whole genomes of lineage B.1.36. The shade of the branches and tips indicates the inferred location state at the nodes. The dark ancestral node of run3/s5 (S5) sequence is annotated with the inferred location and probability. Other sequences collected in Pakistan, but not in this study, are are differentiated by their lighter shade from the ancestral node. Clusters without Pakistan sequences were mostly collapsed to aid visualization of sequences of interest.

24 November 2020 (Figs 2, 3, and 6). A common ancestor that existed between 14 November 2020 and 10 December 2020 in the UK (location probability = 1) produced an introduction that led to a cluster with three sequences, including run3/s3 (S6) with the travel history to the UK and run4/s3 (S20) with a record of the travel history to the UK and layover in Bahrain (Figs 2, 3, and 6). Sequence run3/s4 (S7) had a record of the travel history to the UK and layover in Qatar. We estimated that the viral introduction that resulted in this infection occurred in the UK (location probability = 0.53), likely between 16 October and 29 November 2020 (Figs 2, 3, and 6).

In addition to Fig. 3, a graphical representation of the sequence data available for six cases shown in Fig. 3 includes four cases for B.1.1.7 strains and one case each of B.1.36 and B.1.1.212 (Supplementary Figure S4). The four cases for B.1.1.7 (study subjects S6, S7, S11, and S20) between them had twenty-two household contacts of which eight individuals (36 per cent) were reported to be positive for SARS-CoV-2 (Supplementary Table 4). The one case for B.1.36 (study subject S5) had nine household contacts of which four individuals (44 per cent) were reported to be positive for SARS-CoV-2. Lastly, the one case of B.1.1.212 (study subject S10) had four household contacts of which none were reported to be positive for SARS-CoV-2.

4. Discussion

SARS-CoV-2 variants associated with rapid transmission and increased disease severity—such as the B.1.1.7 (first identified in the UK as an alpha variant), B.1.351 (first identified in South Africa as a beta variant), and P.1 (first identified in Brazil as a gamma variant) variants—have been identified in multiple countries (Drummond et al. 2006; Lauer et al. 2020; CDC 2021; Williams et al. 2021). While investigations into the origin of SARS-CoV-2 variants are ongoing, it is evident that transmission of strains through international travel is of global concern and, wherever possible (Nguyen et al. 2015; Lemey et al. 2020), public health measures such as tracking, tracing, and quarantine of individuals should be carried out to inform countries of introductions of VOCs/Variant of Concern (VOC)/Variants Under Investigation (VUI)/ Variants Of Interest (VOI) and allow them to take appropriate public health actions in response to these variant introductions.

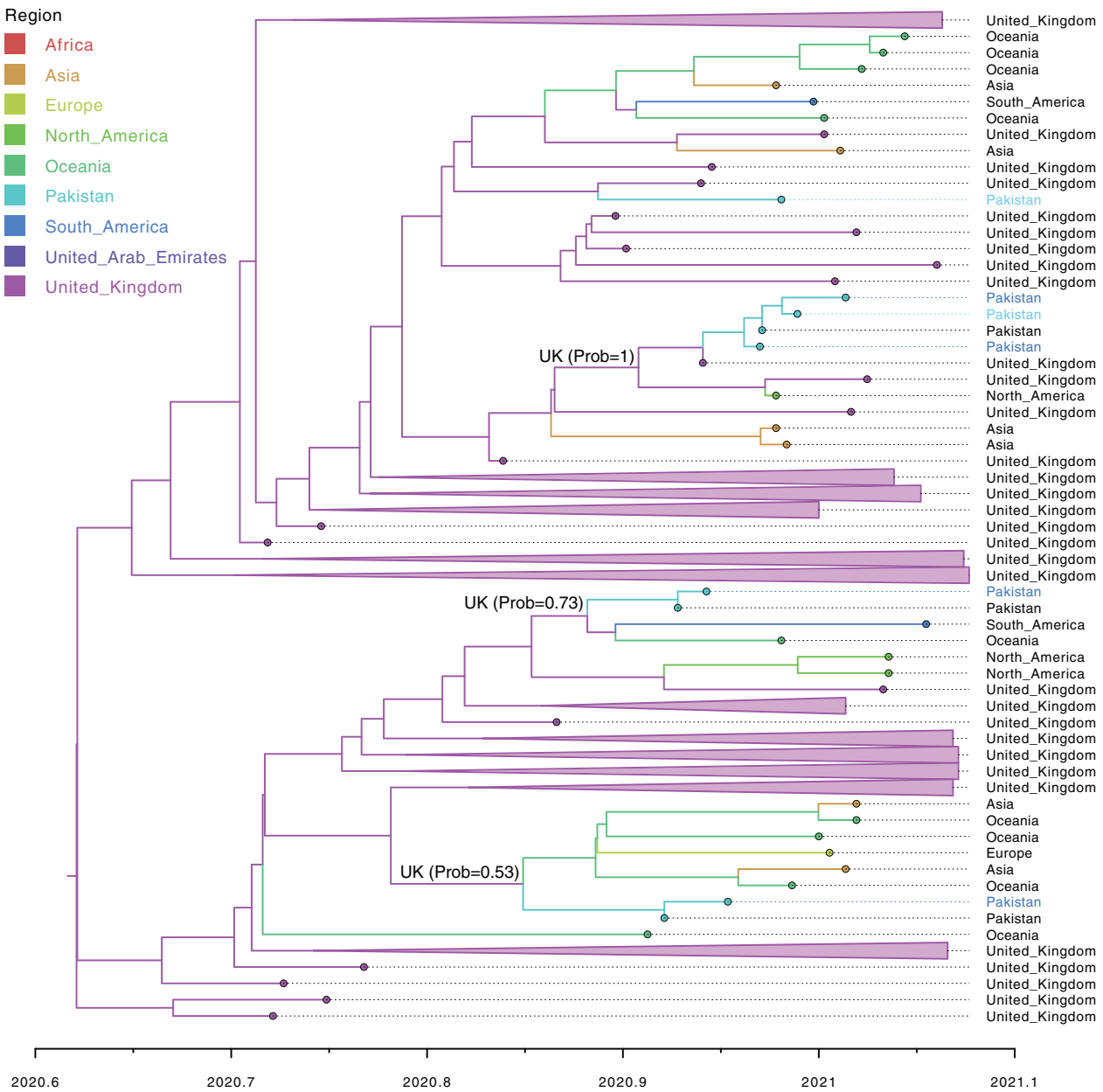

Figure 6. MCC tree inferred for the whole genomes of lineage B.1.1.7. The shade of the branches and tips indicates the inferred location state at the nodes. The ancestral node of run3/s2 (S11), run3/s3 (S6), run3/s4 (S7), and run4/s3 (S20) sequences are annotated with the inferred location and probability. Other sequences collected in Pakistan, but not in this study, are shown by a lighter shade than the ancestral node. Clusters without Pakistan sequences were mostly collapsed to aid visualization of sequences of interest.

The COVID-19 pandemic has demonstrated the value of genomic surveillance of SARS-CoV-2 to supplement traditional epidemiological methods. However, the ability to conduct genomic surveillance has been shown to vary greatly depending on the technical capacity for genomic sequencing and financial resources of different countries. The UK has had one of the largest genomic sequencing efforts through the pandemic; according to the COVID-19 Genomics UK consortium (COG-UK), they have sequenced around 2 million whole SARS-CoV-2 genomes up until March 2022 (https://www.cogconsortium.uk/, last accessed 9 March 2022). However, in developing countries with limited infrastructure for high-level molecular testing, genomic sequencing of such a large number of isolates is not possible due to limitations that include lack of technical expertise and access to equipment, financial costs, and availability of reagents—which became almost inaccessible when global travel and transport restrictions were put in place. Thus, PCR-based approaches with targeted mutation detection provide a cost-effective alternative. PHE introduced the strategy of using the TaqPath (Thermo) assay targeting SGTF-based variation to track the spread of the alpha VOC (PHE, 2021). Reports indicate that >90 per cent of SGTF were B.1.1.7 variants (Nyberg et al. 2021).

We describe the identification of B.1.1.7 lineage and additional SARS-CoV-2 lineages through screening of international travelers between the end of December 2020 and the middle of February 2021. We describe the value of using a combined sequencing and PCR-based approach to identify SARS-CoV-2 variants to guide public health surveillance. We coordinated with the Health Department at the end of December 2020, when there was global concern regarding the global spread of B.1.1.7 variants. The Government of

Pakistan put in place travel restrictions for travelers from countries such as the UK, where the B.1.1.7 strains were present. To limit the introduction of emerging variants at border entry points, Pakistan required that all inbound passengers from the UK must have a negative COVID-19 PCR test within 72 hours prior to travel. Furthermore, it was mandated that all inbound passengers would undergo a SARS-CoV-2 PCR test within 24–48 hours after arrival in Karachi.

We first identified SARS-CoV-2 strains with SGTF using the TaqPath assay to screen for B.1.1.7 lineage isolates. Of thirty-five cases travelers tested, thirteen of these index cases were found to have SARS-CoV-2 SGTF mutants through the TaqPath assay. SGTF mutants were found in both adult and child travelers. All SGTF samples were confirmed to be B.1.1.7 lineage by either sequencing or the PCR-based identification of characteristic B 1.1.7 mutations, i.e., A23063T and C23271A. Genomic sequencing could be conducted for five of these thirteen to confirm them to be B.1.1.7 strains. Two non-SGTF strains were identified through sequencing to be B.1.36 and B.1.1.212 strains. This gives a sense of the heterogeneity of the viral lineages introduced through international travel into Karachi during the given period, end of December 2020 until mid-February 2021 (https://coronavirus.jhu.edu/region/pakistan, last accessed on 9 March 2022).

Contact tracing and testing of COVID-19 cases is an important part of disease control. The 34 index traveler cases studied had 208 household contacts of which 68 were found to be SARS-CoV-2-positive. This indicates an increase of 200 per cent in the number of COVID-19 cases from these individuals.

Samples of all the contacts were not submitted to our laboratory and hence could not be investigated for B.1.1.7 variants. However, based on the available information regarding the number of infected contacts of each traveler, it appears that there was no difference in COVID-19 transmission between cases depending on whether the strains were SGTF or non-SGTF. This may be because transmission of COVID-19 is affected by many parameters other than SARS-CoV-2 variant type such as the strictness of isolation and non-pharmacological interventions which the index case may have followed once at home. The high rate of infections is in concordance with the peak of the second wave in the country between December 2020 and January 2021.

The six genome sequences analyzed phylodynamically belonged to lineages B.1.1.212, B.1.36, and B.1.1.7 and were from individuals with the travel history to the UK, but with layovers in Qatar, Bahrain, and the UAE. We investigated their origins using a recent extension of the Bayesian discrete phylogeographic model that allows the integration of the individual-based travel history. Due to potential bias introduced in the discrete trait reconstruction by the excessive number of sequences from locations around the world, we employed a subsampling strategy that homogeneously selected a specified number of sequences through time and space. We reconstructed that the sequences are the product of at least five independent introductions into Karachi. Patients were estimated to have been infected either in the UK or in their respective layover locations. Despite not otherwise linked, we observed that run3/s3 (S6) and run4/s3 (S20) were likely infected by the same B.1.1.7 viral source, as they group in the same phylogenetic cluster with one other nonstudy sequence. It is also noteworthy that run3/s1 (S10), despite having a record of the travel history to the UK, was inferred to have been the result of an infection acquired in a layover in Bahrain. This makes the present work among the first to apply this novel model, which allowed us to gain insight into the circulating viral diversity in the undersampled country of Bahrain, for which (as of 6 February 2021) there were approximately 150 sequences available. Our study demonstrates the ease with which new respiratory viruses may be introduced through international travel. This reinforces the importance of control measures including SARS-CoV-2 testing prior to travel and/or upon arrival and quarantine requirements.

5. Conclusion

The identification and surveillance of genetic variants of SARS-CoV-2 is of global significance in understanding COVID-19 epidemiology (Fauver et al. 2020; Worobey et al. 2020; Kraemer et al. 2021; Lemey et al. 2021). Genomic sequencing remains prohibitively expensive for low-resource settings with limited technical infrastructure for next-generation sequencing. Targeted PCR-based detection of SARS-CoV-2 mutations can be a rapid, low-cost method of diagnosis. Limitations of this approach include the inability to discover new viral variants and the need to modify assays for new variants. However, PCR-based diagnosis of SARS-CoV-2 VOC/VUI can provide a rapid, effective method to screen for known lineages where genomic sequencing-based surveillance is not possible.

Data availability

All sequence data used in this study are referenced in the Supplementary Table and available upon request.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Acknowledgements

TaqPath reagents were donated for the COVID-19 pandemic response by United Energy Pakistan. We thank the Clinical Laboratories management Sohail Baloch, Naima Maniar, and Nazneen Islam for coordination of COVID-19 screening of travelers. We thank the Health Department, Government of Sindh, for their cooperation in this work. We acknowledge the authors and laboratories that generated and submitted sequences into GISAID's EpiFlu Database. Supplementary Table S2 contains a list of all authors who have contributed to sequences used in this paper. The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Funding

This work received financial support through the Fogarty International Center, National Institutes of Health (NIH), USA; a University Research Council, Aga Khan University grant award; World Health Organization, Pakistan; from Health Security Partners, USA, and Rapid Research Grant—project 236, Higher Education Commission, Pakistan. G.B. acknowledges support from the Internal Fondsen KU Leuven/Internal Funds KU Leuven (Grant No. C14/18/094) and from the Research Foundation—Flanders ("Fonds voor Wetenschappelijk Onderzoek - Vlaanderen," G0E1420N, G098321N).

Conflict of interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Author contributions

A.N., A.B., N.T., U.A. and Z.H.: conceptualization; A.N., A.B., A.K., F.M., N.G., Z.A., M.S., S.S., M.W., Z.A. and U.A.: data curation; A.N., A.B., N.T., A.K. and Z.H.: formal analysis; U.A., R.H., Z.R., D.S. and Z.H.: funding acquisition; A.N., U.B, A.R., N.T., F.M., P.T. and Z.H.: investigation; U.B., A.N., A.R., N.T., P.T., T.M., S.R., M.S. and Z.H.: methodology; U.B., P.T., Z.R. and U.A. Z.H.: project administration; U.A., D.S., Z.R., N.T., R.H., and Z.H.: resources: N.T., B.M., G.B., P.T. and Z.H: software; U.A., G.B., Z.R., R.H., and Z.H.: supervision; A.N., A.K., and Z.H.: validation; visualization: A.N., A.R., N.T., R.H. and Z.H.: roles/writing—original draft; A.N., U.A., A.R., N.T., R.H. and Z.H.: writing—review and editing; A.N., A.R., N.T., Z.R., D.S., U.B., R.H. and Z.H.

Institutional review board statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethical Review Committee, The Aga Khan University, Pakistan.

References

Ayres, D. L. et al. (2012) 'BEAGLE: An Application Programming Interface and High-performance Computing Library for Statistical Phylogenetics', *Systematic Biology*, 61: 170–3.

Bal, A. et al., C. O.-D. H. S. Group. (2021) 'Two-step Strategy for the Identification of SARS-CoV-2 Variant of Concern 202012/01 and Other Variants with Spike Deletion H69-V70, France, August to December 2020', *Euro Surveillance*, 26: 2100008.

CDC. (2021) 'Science brief: Emerging SARS-CoV-2 Variants'.

Challen, R. et al. (2021) 'Risk of Mortality in Patients Infected with SARS-CoV-2 Variant of Concern 202012/1: Matched Cohort Study', *BMJ*, 372: n579.

Drummond, A. J. et al. (2006) 'Relaxed Phylogenetics and Dating with Confidence', *PLoS Biology*, 4: e88.

Edwards, C. J. et al. (2011) 'Ancient Hybridization and an Irish Origin for the Modern Polar Bear Matriline', *Current Biology: CB*, 21: 1251–8.

Elbe, S., and Buckland-Merrett, G. (2017) 'Data, Disease and Diplomacy: GISAID's Innovative Contribution to Global Health', *Global Challenges*, 1: 33–46.

Fauver, J. R. et al. (2020) 'Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States', *Cell*, 181: 990–6 e995.

Ferguson, N. M. et al. (2020) 'Report 9: Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand. 2020'. Imperial College, London, UK.

Frampton, D. et al. (2021) 'Genomic Characteristics and Clinical Effect of the Emergent SARS-CoV-2 B.1.1.7 Lineage in London, UK: A Whole-genome Sequencing and Hospital-based Cohort Study', *The Lancet Infectious Diseases*, 21: 1246–56.

Ghanchi, N. K. et al. (2021) 'Higher Entropy Observed in SARS-CoV-2 Genomes from the First COVID-19 Wave in Pakistan', *PLoS One*, 16: e0256451.

Huang, C. et al. (2020) 'Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China', *The Lancet*, 395: 497–506.

Imran, M. et al. (2021) 'COVID-19 Situation in Pakistan: A Broad Overview', *Respirology*, 26: 891–2.

JHU. (2021) 'COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University'.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019) 'MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization', *Briefings in Bioinformatics*, 20: 1160–6.

Kraemer, M. U. G. et al. (2021) 'Spatiotemporal Invasion Dynamics of SARS-CoV-2 Lineage B.1.1.7 Emergence', *Science*, 373: 889–95.

Lauer, S. A. et al. (2020) 'The Incubation Period of Coronavirus Disease 2019 (COVID-19) from Publicly Reported Confirmed Cases: Estimation and Application', *Annals of Internal Medicine*, 172: 577–82.

Lemey, P. et al. (2020) 'Accommodating Individual Travel History and Unsampled Diversity in Bayesian Phylogeographic Inference of SARS-CoV-2', *Nature Communications*, 11: 5110.

——— et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5: e1000520.

——— et al. (2021) 'Untangling Introductions and Persistence in COVID-19 Resurgence in Europe', *Nature*, 595: 713–7.

Nguyen, L. T. et al. (2015) 'IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-likelihood Phylogenies', *Molecular Biology and Evolution*, 32: 268–74.

Nyberg, T. et al. (2021) 'Risk of Hospital Admission for Patients with SARS-CoV-2 Variant B.1.1.7: Cohort Analysis', *BMJ*, 373: n1412.

Pakistan, G. O. (2021) 'COVID-19 Health Advisory Platform. Ministry of National Health Services Regulations and Coordination'.

PHE. 'Investigation of Novel SARS-CoV-2 Variant of Concern 2020/12/01Technical Briefing 4. 2021, Public Health England'.

Plante, J. A. et al. (2021) 'Spike Mutation D614G Alters SARS-CoV-2 Fitness', *Nature*, 592: 116–21.

Quick, J. N. (2019). 'CoV-2019 Sequencing Protocol V3 (Locost) V.3. 2020'.

Rambaut, A. et al. (2018) 'Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7', *Systematic Biology*, 67: 901–4.

——— et al. (2020) 'A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology', *Nature Microbiology*, 5: 1403–7.

——— et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: vew007.

Suchard, M. A. et al. (2018) 'Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10', *Virus Evolution*, 4: vey016.

Tamim, S. et al. (2021) 'Genetic and Evolutionary Analysis of SARS-CoV-2 Circulating in the Region Surrounding Islamabad, Pakistan', *Infection, Genetics and Evolution*, 94: 105003.

Tavaré, S. (1985) 'Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences', *Lectures on Mathematics in the Life Science*, 16: 57–58.

Toole, O. A. et al. (2021) 'Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pan-golin Tool', *Virus Evolution*, 7: veab064.

Umair, M. et al. (2021) 'Whole-genome Sequencing of SARS-CoV-2 Reveals the Detection of G614 Variant in Pakistan', *PLoS One*, 16: e0248371.

Williams, H., Hutchinson, D., and Stone, H. (2021) 'Watching Brief: The Evolution and Impact of COVID-19 Variants B. 1.1. 7, B. 1.351, P. 1 And B. 1.617', *Global Biosecurity*, 3.

Worobey, M. et al. (2020) 'The Emergence of SARS-CoV-2 in Europe and North America', *Science*, 370: 564–70.

Wu, F. et al. (2020) 'A New Coronavirus Associated with Human Respiratory Disease in China', *Nature*, 579: 265–9.

Yang, Z. (1994) 'Maximum Likelihood Phylogenetic Estimation from DNA Sequences with Variable Rates over Sites: Approximate Methods', *Journal of Molecular Evolution*, 39: 306–14.