

RESEARCH

Open Access



Sequential viral introductions and spread of BA.1 across Pakistan provinces during the Omicron wave

Ali Raza Bukhari¹, Javaria Ashraf¹, Akbar Kanji¹, Yusra Abdul Rahman¹, Nídia S. Trovão², Peter M. Thielen³, Maliha Yameen¹, Samiah Kanwar⁴, Waqasuddin Khan⁴, Furqan Kabir⁴, Muhammad Imran Nisar⁴, Brian Merritt³, Rumina Hasan¹, David Spiro², Zeba Rasmussen², Uzma Bashir Aamir⁵ and Zahra Hasan^{1*}

Abstract

Background COVID-19 waves caused by specific SARS-CoV-2 variants have occurred globally at different times. We focused on Omicron variants to understand the genomic diversity and phylogenetic relatedness of SARS-CoV-2 strains in various regions of Pakistan.

Methods We studied 276,525 COVID-19 cases and 1,031 genomes sequenced from December 2021 to August 2022. Sequences were analyzed and visualized using phylogenetic trees.

Results The highest case numbers and deaths were recorded in Sindh and Punjab, the most populous provinces in Pakistan. Omicron variants comprised 93% of all genomes, with BA.2 (32.6%) and BA.5 (38.4%) predominating. The first Omicron wave was associated with the sequential identification of BA.1 in Sindh, then Islamabad Capital Territory, Punjab, Khyber Pakhtunkhwa (KP), Azad Jammu Kashmir (AJK), Gilgit-Baltistan (GB) and Balochistan. Phylogenetic analysis revealed Sindh to be the source of BA.1 and BA.2 introductions into Punjab and Balochistan during early 2022. BA.4 was first introduced in AJK and BA.5 in Punjab. Most recent common ancestor (MRCA) analysis revealed relatedness between the earliest BA.1 genome from Sindh with Balochistan, AJK, Punjab and ICT, and that of first BA.1 from Punjab with strains from KPK and GB.

Conclusions Phylogenetic analysis provides insights into the introduction and transmission dynamics of the Omicron variant in Pakistan, identifying Sindh as a hotspot for viral dissemination. Such data linked with public health efforts can help limit surges of new infections.

Keywords Omicron, Variants of concern, SARS-CoV-2, Phylogenetics, Transmission dynamics

*Correspondence:

Zahra Hasan
zahra.hasan@aku.edu

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



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

The rise and fall of coronavirus disease 2019 (COVID-19) cases since the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in December 2019 has driven the need for monitoring of variants through genomic surveillance [1]. The global burden of COVID-19 is not fully known although greater than 754 million cases have been reported as of February 3, 2022. Pakistan has reported about 1.5 million cases of COVID-19 and nearly 31,000 deaths [2, 3]. COVID-19 vaccines have had a significant impact on controlling both morbidity and mortality from COVID-19 [4, 5]. The proportion of deaths to cases changed greatly throughout the pandemic with reduced disease severity after vaccines were introduced and the evolution of SARS-CoV-2 variants of concern (VoC) toward higher transmissibility and lower pathogenicity. The epidemiology of COVID-19 has been informed by the reported number of cases, deaths, hospitalizations, and viral genomic sequencing. This information has varied greatly between high- and low- income countries of similar population sizes; explanatory factors include lack of resources needed for gathering metadata, PCR testing and genomic sequencing [6].

SARS-CoV-2 VoCs have shown a trend toward increased transmissibility, leading to strain-specific increases in COVID-19 cases [7]. The first VoC was the Alpha variant or B.1.1.7, followed by Beta/B.1.351 and Delta/B.1.617.2 strains in 2021. VoCs which emerged after the introduction of vaccinations were more effective at evading host immunity driven by both natural infection and vaccinations [8]. This was evidenced by Omicron and its subvariants coming to dominate over other lineages across the globe by the end of 2021 [9, 10]. Omicron has distinct subvariants, five of which include BA.1, BA.2, BA.3, BA.4 and BA.5. BA.1 was first identified in South Africa and Botswana on November 26, 2021 [11]. By January 2022, BA.1 made up 81% of cases in South Africa, rapidly decreasing to 45% in February 2022. BA.5 was first identified in South Africa on January 21, 2022, and spread elsewhere, particularly in Europe, by April 2022. Soon after this time, BA.4 and BA.5 variants became predominant [12].

Understanding the relationship between COVID-19 rates and SARS-CoV-2 variants at the regional level is important for management of the public health response. The identification of new SARS-CoV-2 variants and their prevalence in different regions can help make public health strategies, such as vaccination campaigns, testing protocols, and contact tracing efforts. By monitoring the circulation of SARS-CoV-2 variants and their impact on COVID-19 rates, public health officials can make decisions about the allocation of resources and

implementation of preventive measures to limit the spread of the virus.

Pakistan has a population of approximately 220 million people. It consists of seven distinct regional territories: Azad Jammu and Kashmir (AJK), Balochistan, Gilgit-Baltistan (GB), Khyber Pakhtunkhwa (KP), Punjab, Sindh and the Islamabad Capital Territory (ICT). The first COVID-19 case from Pakistan was reported in Sindh on February 26, 2020. The timeline of pandemic waves experienced in the country was; March to July 2020 where the G Nextstrain clade of SARS-CoV-2 dominated, October 2020 to January 2021 dominated by GR/GH clades, April to May 2021 where Alpha dominated [13], July to September 2021 where Delta was the predominant strain, and, December 2021 to February 2022 dominated by Omicron.

Here, we studied COVID-19 waves in association with cases and mortality and Omicron subvariants identified in regions across Pakistan. We also investigated the introduction of SARS-CoV-2 strains and studied the relatedness of variants in each region to understand their patterns of transmission.

Materials and methods

This study was approved by the Ethical Review Committee, The Aga Khan University (AKU), Pakistan.

Data used in this study

The COVID-19 case numbers presented in this study were obtained from the John Hopkins Coronavirus resource center which collects data from the official website of each country as mentioned in the article by Dong et al. [14]. As per the Github repository <https://github.com/CSSEGISandData/COVID-19> the source data for Pakistan is the Pakistan Government official COVID-19 website, <http://covid.gov.pk/stats/pakistan>. Thus, the numbers from the JHU Coronavirus resource center reflect the aggregate of COVID-19 positive cases reported from regional laboratories across Pakistan.

Selection of SARS-CoV-2 genomes

The genomes included in this study were those obtained from clinical isolates identified from respiratory specimens received for SARS-CoV-2 diagnostic testing at the Aga Khan University (AKU) Hospital Laboratories and other laboratories in Pakistan. SARS-CoV-2 strains sequenced included but were not limited to those conducted at AKU, Karachi [15], National Institute of Health, Islamabad [16] and other laboratories thereby representing nationwide data. AKU contributed to 324 of the 1031 genomes analysed here, Supplementary Table 1. Samples collected at AKU were from Sindh province and were screened on a regular basis whereby up to five

SARS-CoV-2 PCR positive samples were collected each day and those which met sequencing criteria of a CT value ≤ 26 were processed for whole genome sequencing (WGS). Overall, the data from GISAID represents samples collected as a part of SARS-CoV-2 genomic surveillance across Pakistan and were identified based on collection dates between December 1, 2021 and August 14, 2022. The sequenced were downloaded from GISAID on September 15, 2022 by selecting those submitted from Pakistan and which were “complete”, whilst, “low coverage excluded” was part of the filtering criteria. In total, we downloaded 1031 sequences with metadata; of these, 955 were found to belong to the Omicron VoC, of which 952 had complete metadata and were used for phylogenetic analysis.

COVID-19 cases and mortality data

COVID-19 case data for the period December 1, 2021 to August 14, 2022 were downloaded from the John Hopkins coronavirus resource center [3] and used for analysis. COVID-19 mortality data were downloaded from Pakistan’s official COVID-19 page (<https://covid.gov.pk/>) accessed on September 1, 2022.

Phylogenetic tree and analysis

FASTA files of the 952 Omicron sequences with corresponding metadata (age, gender, date of collection and location) were used for phylogenetic tree reconstruction using the augur pipeline [17]. Out of the total (952), 944 genomic sequences qualified for phylodynamic mapping.

Full length SARS-CoV-2 genomes of Omicron subvariants were aligned using the MAFFT alignment tool [18]. Multiple Sequence Alignment (MSA) files generated from MAFFT were used for a maximum likelihood (ML) phylogenetic tree through IQ-TREE2 [19]. By applying a generalized midpoint rooting strategy, rooting of the tree was carried out with branch length variance using Tree-Time [20]. The tree was visualized and edited in Figtree v. 1.4.4 (<https://tree.bio.ed.ac.uk>). The final tree with annotated nodes and metadata was exported to the phylodynamic visualizing tool Auspice [21].

Statistical analysis

Demographic results are presented in mean \pm SD. Kruskal–Wallis statistical tests were used to analyze statistical significance, with p-value less than 0.05 considered statistically significant. Graph Pad Prism v. 5.0 (<http://www.graphpad.com>) was used for statistical analysis.

Results

Infections and deaths in different regions of Pakistan during the fifth COVID-19 wave

A total of 276,525 cases were reported between December 2021 and August 2022 [22]. The region-wide distribution of COVID-19 cases is depicted in Fig. 1, which also depicts the population density of each of the 7 regions studied. Sindh reported the most cases (41.7%; 2413 persons per million), followed by Punjab with 27.1% (683 persons per million), 15.2% from KP (1185 persons per million), 11% from ICT (15199 persons per million),

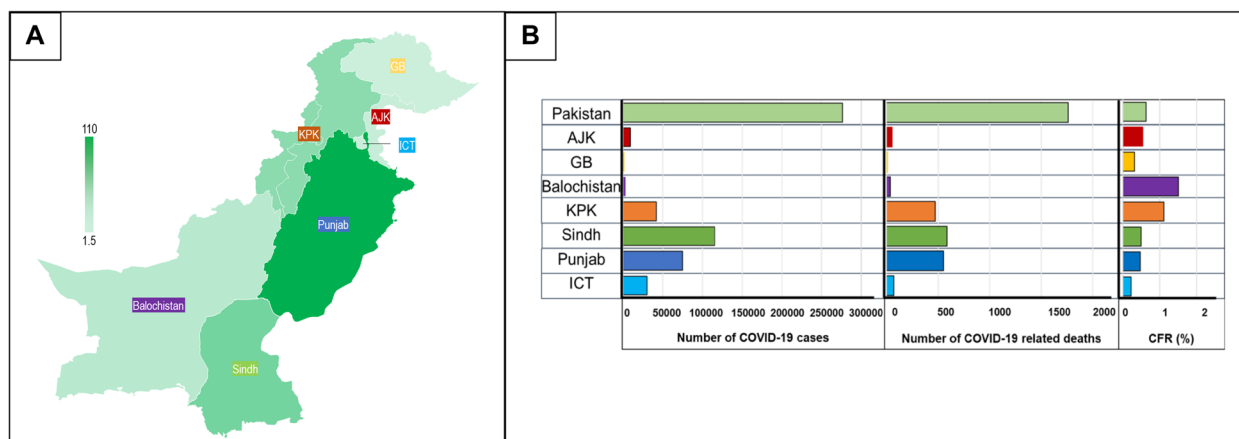


Fig. 1 Population density, COVID -19 cases, deaths, and CFR across Pakistan. The graphs depict nationwide data from December 1, 2021 until August 14, 2022. **A** The region-wise population density of Azad Jammu and Kashmir (AJK), Balochistan, Gilgit-Baltistan (GB), Islamabad Capital Territory (ICT), Khyber Pakhtunkhwa (KPK), Punjab and Sindh is presented. The scale bar displays population values in millions of persons shaded by color. **B** Left panel x-axis shows number of COVID-19 cases, the middle panel shows number of COVID-19 related deaths, and the right panel shows the case fatality rate (CFR %) per regions colored as: ICT (light blue), Punjab (dark blue), Sindh (green), KPK (orange), Balochistan (purple), GB (yellow) and AJK (red)

3.4% from AJK (2338 persons per million), 0.9% from Balochistan (193 persons per million) and 0.6% from GB (1040 persons per million) of total cases.

A total of 1,766 COVID-19 deaths were reported during the study period (Fig. 1). Sindh reported the greatest (33.1% of total deaths) followed by Punjab (31.1%), KP (26.7%), ICT (3.9%), AJK (2.9%), Balochistan (2.0%) and GB (0.3%). The overall case fatality ratio percentage (CFR%) over this period was 0.6%. It was highest in Balochistan (1.5%), followed by KP (1.1%), AJK (0.5%), Sindh (0.5%), Punjab (0.5%), GB (0.3%) and ICT (0.2%).

COVID-19 case numbers rose rapidly from under 1,000 to more than 10,000 per week across Pakistani regions between the end of January and beginning of February 2022, with 22,000 cases being reported in the last week of January 2022 in Sindh alone (Fig. 2). COVID-19 peaks occurred sequentially in other provinces; ICT by the last week of January 2022; Punjab, KP, Balochistan and AJK by the first week of February 2022; and GB by the second week of February 2022. Cases declined around early March 2022, reaching a few hundred cases per week by June 2022. Another slight rise in cases was observed in Sindh (early July 2022) and Punjab (early August 2022). A similar trend was also observed in other regions. Of note, a rise in COVID-19 cases was observed in Sindh ahead of the other regions of Pakistan.

The mean age of COVID-19 cases across Pakistan was 39 SD ± 19 years. For each region the mean age was: ICT 38 SD ± 19, Punjab 39 SD ± 21, Sindh 40 SD ± 19, KP 32 SD ± 19, Balochistan 48 SD ± 17, AJK 38 SD ± 19 and GB 43 SD ± 24 years. There was no significant difference between age groups of COVID-19 cases across the regions.

Heterogeneity in data submission across regions of Pakistan

We investigated the association of COVID-19 waves with SARS-CoV-2 Omicron variants. We performed phylogenetic analysis of the 957 Omicron genomes available in relation to their date and location of sample collection. The monthly rate of submissions across the study period was not uniform (Supplementary Fig. 1 and 2). More genomes were submitted in March, June, July and August 2022. There was variability in the location of the submissions. Provinces with laboratories with genomic surveillance capacity had greater representation; Sindh (n=364), ICT (376) and Punjab (n=117) contributed more than 80% of total submissions, with limited representation from GB (n=43), KP (n=61), and AJK (n=63). The fewest SARS-CoV-2 genomes were from Balochistan (n=14).

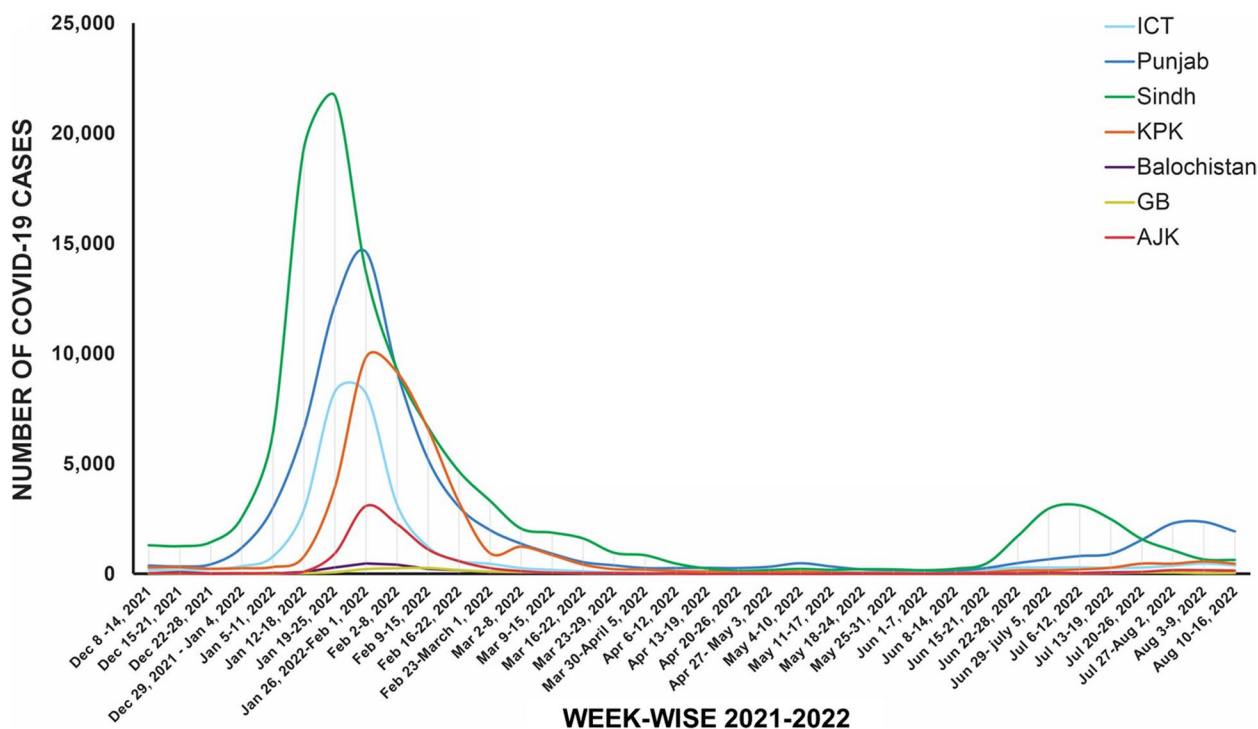


Fig. 2 Trend of COVID-19 cases in Pakistan. The figure depicts the weekly count of COVID-19 cases through the period from December 1, 2021, until August 14, 2022. Cases are shown region-wise; ICT (light blue), Punjab (dark blue), Sindh (green), Khyber Pakhtunkhwa (KPK, orange), Balochistan (purple), GB (yellow) and AJK (red)

The pango lineage distribution was A (0.3%), AY (4.8%), B (1.8%), B.1 (0.8%), BA.1 (Nextstrain 21 K; 16.6%), BA.2 (Nextstrain clade 21L; 31.5%), BA.4 (Nextstrain 22A; 3.6%) and BA.5 (Nextstrain 22B; 40.6%).

We also investigated the association between variants and age of COVID-19 cases; we divided the data for sequences available into 4 groups (≤ 18 years, 19–40 years, 41–55 years and ≥ 56 years). For each of the SARS-CoV-2 variants, we found the greatest number of cases to be in those aged 19–40 years, $p < 0.0001$ (Table 1).

Omicron and subvariants across regions of Pakistan

The frequency of Omicron subvariant sequences observed across different regions of Pakistan is depicted in Fig. 3. ICT uploaded 376 sequences to GISAID, the greatest of which were Omicron mainly, BA.5 ($n = 153$, 43.6%) and BA.2 ($n = 153$, 43.6%). There were 117 sequences from Punjab, predominantly BA.5 ($n = 45$, 41.3%) and BA.2 ($n = 43$, 39.4%). The distribution of 354 genomes from Sindh included BA.5 ($n = 135$, 41.9%), BA.2 ($n = 76$, 23.6%) and BA.1 ($n = 102$, 31.7%). The fewest genomes were submitted from AJK ($n = 63$, 6%), KP ($n = 61$, 6%) and GB ($n = 43$, 4%); of those available,

Table 1 Distribution of SARS-CoV-2 lineages according to different age groups

SARS-CoV-2 variant (number, %)	Age (years)	number of variants in each group	% of variant in each group (%)	p-value (Kruskal-Wallis)
A ($n = 3$, 0.3%)	≤ 18	0	0.0	not possible
	19–40	2	66.6	
	41–55	1	33.3	
	≥ 56	0	0.0	
AY ($n = 46$, 4.7%)	≤ 18	3	6.5	< 0.0001
	19–40	29	63.0	
	41–55	8	17.4	
	≥ 56	6	13.0	
B ($n = 17$, 1.8%)	≤ 18	7	41.2	0.0029
	19–40	7	41.2	
	41–55	2	11.8	
	≥ 56	1	5.9	
B.1 ($n = 8$, 0.8%)	≤ 18	3	37.5	0.0933
	19–40	3	37.5	
	41–55	1	12.5	
	≥ 56	1	12.5	
BA.1 ($n = 164$, 16.9%)	≤ 18	16	9.75	< 0.0001
	19–40	71	43.2	
	41–55	30	18.2	
	≥ 56	47	28.6	
BA.2 ($n = 305$, 31.5%)	≤ 18	36	11.8	< 0.0001
	19–40	146	47.9	
	41–55	57	18.7	
	≥ 56	66	21.6	
BA.4 ($n = 35$, 3.6%)	≤ 18	5	14.3	< 0.0001
	19–40	16	45.7	
	41–55	8	22.9	
	≥ 56	6	17.1	
BA.5 ($n = 392$, 40.4%)	≤ 18	41	10.5	< 0.0001
	19–40	199	50.8	
	41–55	68	17.3	
	≥ 56	84	21.4	

The age-group wise SARS-CoV-2 variants frequency is presented. The Kruskal-Wallis test was used to study association between the age groups, p-value ≤ 0.05 was considered significant

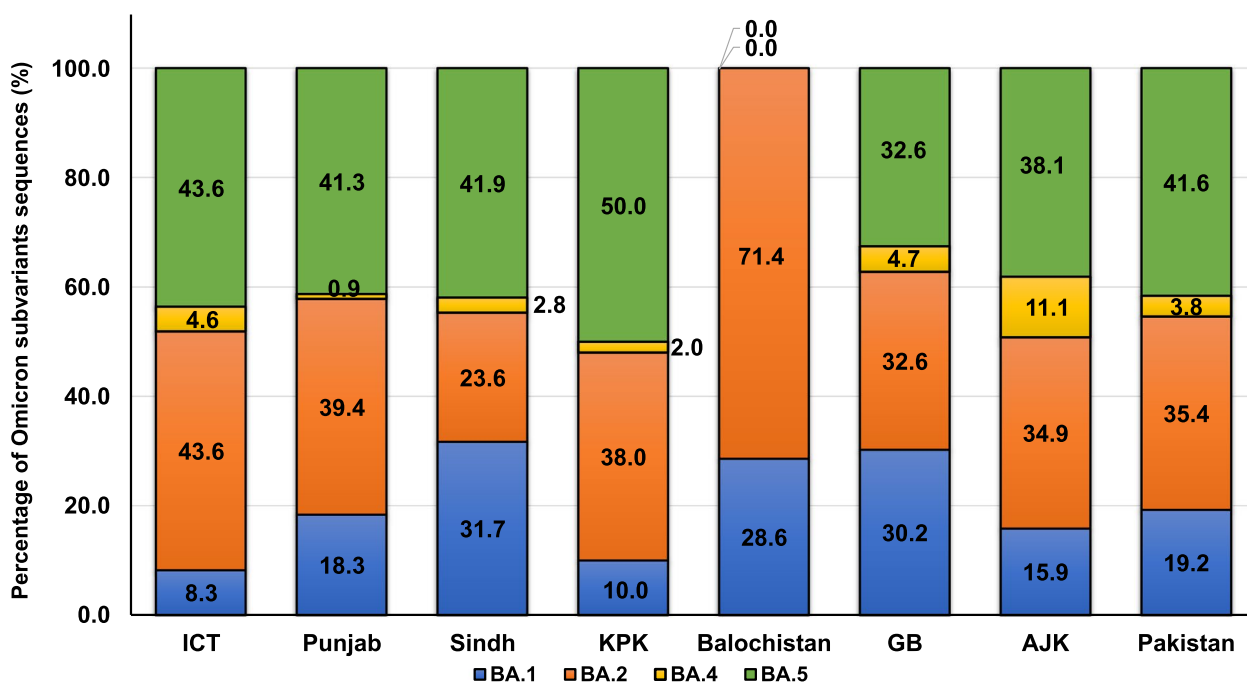


Fig. 3 Frequency of Omicron variants across Pakistan. The graph shows all Omicron (n=955) genomes submitted from each region between December 1, 2021, and August 14, 2022. BA.1 (blue), BA.2 (orange), BA.4 (yellow) and BA.5 (green)

BA.2 and BA.5 were the predominant subvariants. The fewest sequences were from Balochistan ($n=14$, 2%), which did not report any BA.5 subvariant.

Phylogenetic Analysis of SARS-CoV-2 Omicron variants from Pakistan

The phylogenetic analysis of the Omicron genomes across Pakistan is depicted in Fig. 4. The phylogram presents the evolution and spread over time of BA.1, BA.2, BA.4 and

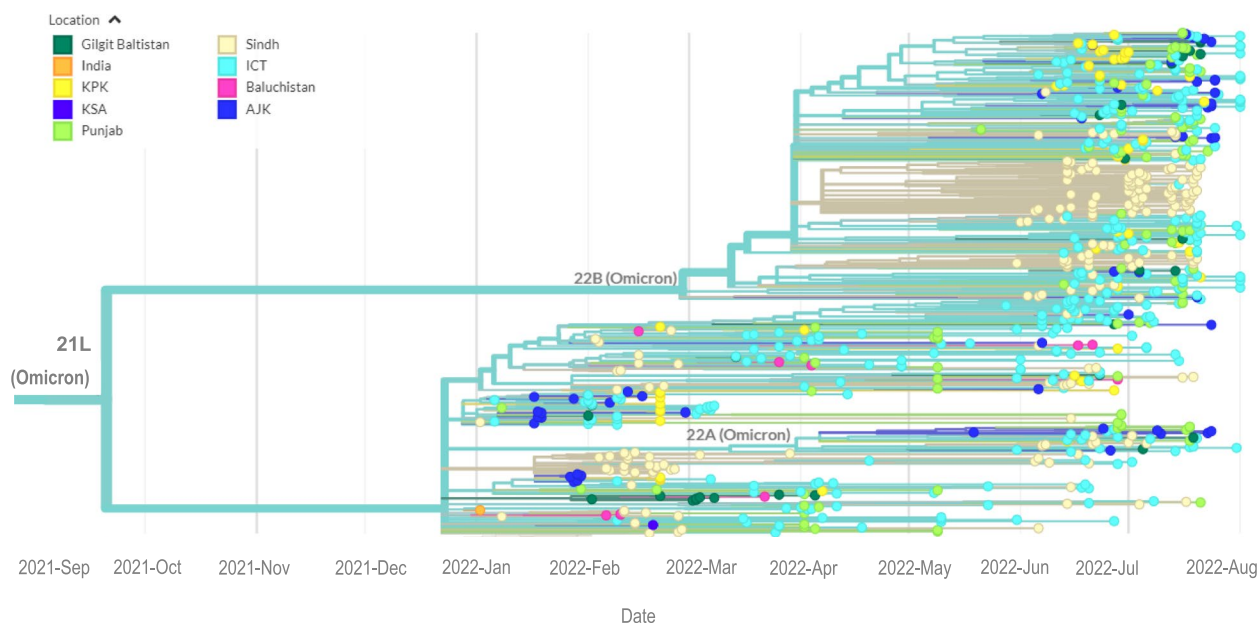


Fig. 4 Phylogenetics of early Omicron variants in Pakistan. The tree illustrates the relatedness of 944 omicron sequences submitted from Pakistan between December 1, 2021, and August 14, 2022. The tree uses color coding to identify the travel origin of each case reported

BA.5 variants across the country including sequences from ICT, Sindh, Punjab, AJK, KP, GB and Balochistan. Some were from travelers who entered Sindh from the Kingdom of Saudi Arabia (KSA, $n=3$), the United Arab Emirates (UAE, $n=2$), India ($n=1$), Turkey ($n=1$) and the United States of America (USA, $n=1$).

Investigating the phylogeny of BA.1, BA.2, BA.4 and BA.5 subvariants

The first BA.1 subvariant was detected on December 13, 2021, and was followed by a surge of cases. Subsequent surges were associated with the BA.2 subvariant and then the BA.4 and BA.5 subvariants. To understand this trend, we separately looked at the phylogenetics of each variant.

The first case of BA.1 was identified in Sindh (Fig. 5A). The BA.1 lineage encompassed sub-lineages BA.1.1, BA.1.1.1, BA.1.1.13, BA.1.1.14, BA.1.1.18, BA.1.15, BA.1.15.1, BA.1.17, and BA.1.18. Subsequently, BA.1 was reported in Punjab, ICT, and AJK. Later, reports of BA.1 were obtained from KP (January 2022) and Balochistan (February 2022).

We next examined the phylogeny of the 317 BA.2 sequences found between January and July 2022. The first BA.2 subvariant was reported in January 2022 from

Sindh, and clustered closely with an isolate identified in a traveler from India (Fig. 5B). Subsequently, BA.2 was reported in ICT and Punjab in April 2022.

The BA.4 subvariant was first reported in AJK in May 2022, followed by reports in June 2022 onwards from ICT, Sindh, KP, Punjab and GB (Fig. 5C).

The BA.5 subvariant was first reported in Punjab in May 2022, followed by reports from Sindh, ICT, AJK, KP and Gilgit (Fig. 5D).

Phylogenetic relatedness of cases reported across different regions of Pakistan

To further understand the genetic relatedness of Omicron subvariants across Pakistan, we analyzed the relatedness of sequenced genomes through the inference of the most recent common ancestor (MRCA) across all 944 genomes (Fig. 6A), focusing on the first BA.1 case reported in each region. The first BA.1 case from Balochistan, with a sampling date of January 21, 2022, shared a common ancestral node with one reported earlier from Sindh on January 10, 2022 (Fig. 6B). The first case from Sindh, reported on December 8, 2022, had no known association with an overseas traveler [23] (Fig. 6C). The first BA.1 case from AJK (January 8, 2022) shared

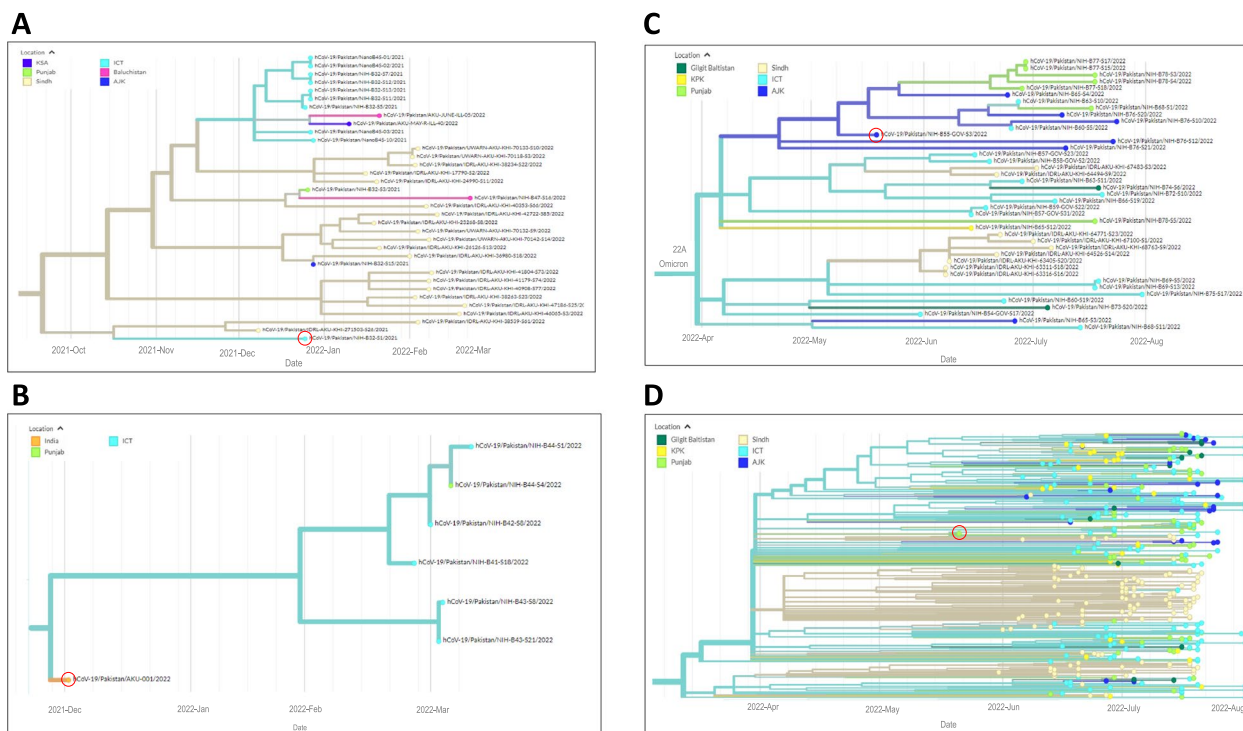


Fig. 5 Introduction and linkage of omicron variants in Pakistan. Phylogenetic trees depict the first case report for each variant as a red circle in the identified timeline with relatedness to later isolates. Panels A-D depict the first case of each variant identified by a red circle for A, BA.1; B, BA.2; C, BA.4 and D, BA.5. Regional locations are identified by colors; Sindh in light yellow, Balochistan in pink, AJK in blue, Punjab in light green, GB in dark green and ICT, light blue. Data are presented as auspice output of the tree generated using IQ-TREE v. 2.2.0

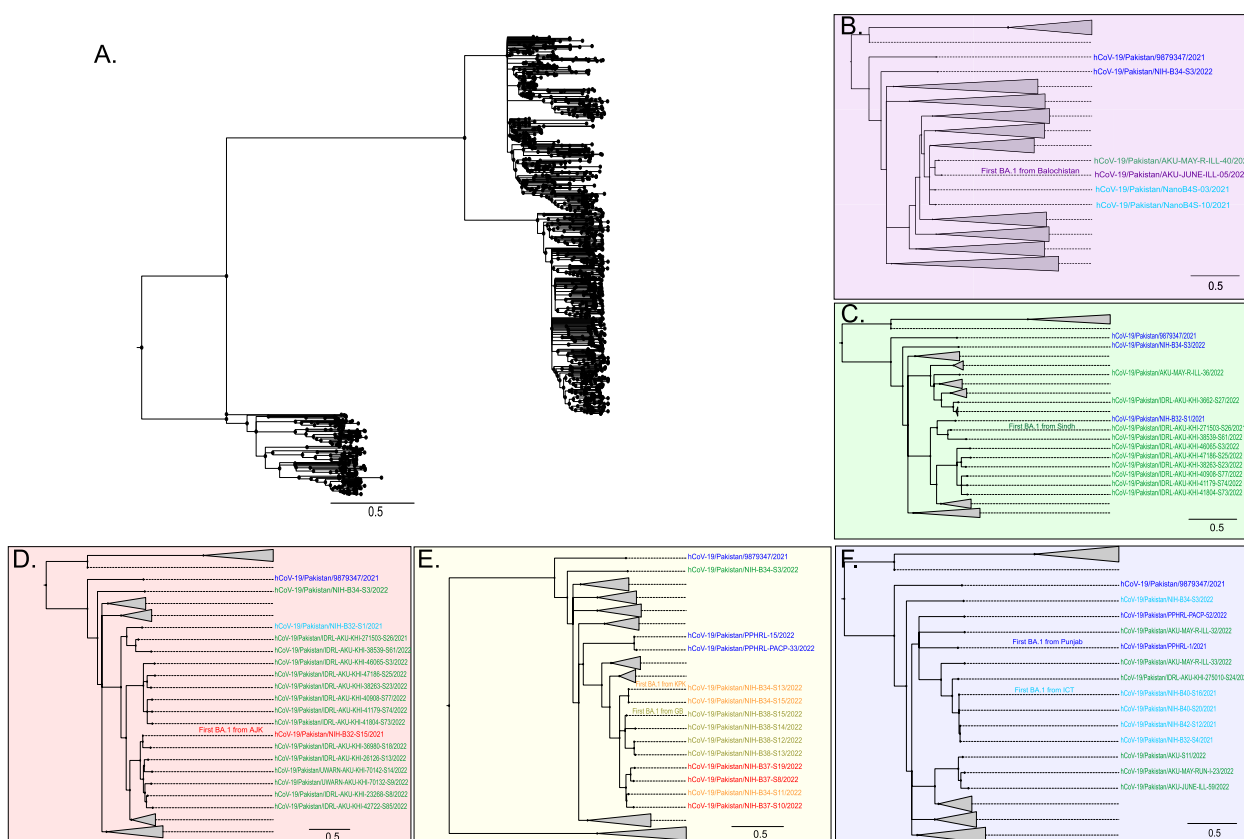


Fig. 6 Evolutionary relatedness of the first reported Omicron variant from each region of Pakistan. **A** Phylogenetic tree for 944 Omicron genomes and related strains sharing the most recent common ancestor (MRCA). The first BA.1 sequence from each region, **B** Balochistan (purple); **C** Sindh (green); **D** AJK (light red); **E** Gilgit-Baltistan and KPK (yellow); **F** Punjab and ICT (light blue). Data are presented as Figtree v. 1.4.4 outputs using the maximum-likelihood tree generated using IQTREE v. 2.2.0. Grey triangles represent collapsed nodes of relatively distant sequences from the first Omicron sequences in each region. Scale bar presents nucleotide substitution/site

an ancestral node with a case from Sindh and Punjab (Fig. 6D). The earliest case from GB and KP had the same MRCA and shared an ancestral node with isolates from the Punjab (Fig. 6E). Similarly, the earliest BA.1 cases from ICT and Punjab shared an ancestral node with Sindh (Fig. 6F).

Discussion

Genomic surveillance data from Pakistan has been limited especially, for early pandemic waves of 2020 and 2021. Our study provides insights into the phylogenetic relatedness of different Omicron variants which spread across Pakistan, associating the introduction of the Omicron subvariants with COVID-19 surges in different regions. This is the first study highlighting the genetic relatedness of Omicron subvariants in Pakistan through an analysis of the MRCAs. We identify Sindh as a hotspot for variant introductions into the country.

COVID-19 waves displayed sequential geographic transmission, occurring first in Sindh, followed by Punjab and other regions. Whilst its population size is less than

half of Punjab’s, Sindh reported 41.7% of all COVID-19 cases. Balochistan reported 0.9% of total cases whilst having the smallest population of all the regions. However, the CFR in Balochistan was the highest (1.5%) among all regions, indicating that the pandemic’s impact was severe there. In a similar manner, KP has a population of 36 million, but has much higher deaths as compared to Sindh (48 million population) and Punjab (110 million population), as reflected by the CFR (1.1%).

Overall, the majority of COVID-19 cases were aged between 19 and 40 years. Deaths have previously been associated with older age groups [24]. The lack of a provincial sequencing facility in Balochistan limits available sequencing efforts and the fewest genomes were submitted from the province. Another challenge here was the limited access to healthcare facilities in Balochistan, a large but sparsely populated province (comprising 6% of population with 43.6% of the land area of Pakistan). One study showed a COVID-19 positivity of 13% between March and December 2020, with the mean age of positive cases to be 36 ± 14 years, with 20% of the cases being

female [25]. Hence, due to the limited data available it is not possible to understand fully the trend of COVID-19 in Balochistan.

Separately, over the course of the pandemic, the highest COVID-19 mortality (3.5%) reported in 2020 was from Peshawar in KP province. Local experts have suggested the lack of social distancing and a non-compliant response to standard operating procedures (SOPs) in the community to be the major reason behind this [26]. Other factors potentially contributing to the higher CFR could include a delayed presentation of COVID-19 symptoms combined with limited testing and reporting.

The first introduction was BA.1 on 8th December 2021 into Karachi, Sindh. BA.1 was followed by reports from Punjab, ICT and AJK. BA.1 reports from Balochistan and KP occurred after January 2022. Soon after identification of BA.1 strains, Pakistan experienced a surge of cases across the country, with particularly high numbers of cases and associated deaths in Karachi, Sindh. The identification of a larger than average case count in the Sindh region was associated with travelers from other destinations. Earlier, the introduction of the Alpha VoC was associated with international travelers to Karachi [27].

The first case from Sindh appeared to have transmitted locally as reported in previous studies [23]. The relatedness of strains between provinces was evident such that, the earliest BA.1 cases from ICT and Punjab share an ancestral node with Sindh. The first BA.1 strain from Balochistan was related to that from Sindh. The first BA.1 case from AJK shared an ancestral node with a case from Sindh and Punjab. The earliest case from GB and KP had the same MRCA and shared an ancestral node with isolates from the Punjab. BA.1 cases in GB, KP, and AJK regions shared a same ancestral node, which might be due to the proximity of these three provinces in the northern region of Pakistan. The spread of variants could be through travel between these regions, but could also be attributed to an independent introduction from elsewhere. The provinces have a shared border with frequent routine travel between them, allowing easy spread, however phylogeographic analysis would be needed to test this hypothesis.

The first BA.2 variant was reported in January 2022 from Karachi, Sindh, and clustered closely with the isolate identified from a traveler from India. BA.2 variants were subsequently reported in ICT and then Punjab. BA.4 was first reported from AJK (May 2022) and BA.5 from Punjab (May 2022). The reporting of newly introduced variants was followed by local transmission. This was evident from reports in the same city/location in addition to those from other provinces.

It is likely that the rise in cases observed in Sindh (July 2022), Punjab (August 2022) and similar trends observed

in other regions was due to the spread of BA.4 and BA.5 strains at this time.

The introduction of Omicron variants in Karachi, Sindh, followed by a rise in COVID-19 cases associated with each wave may be due to the characteristics of the location. The high case count in Sindh was likely driven by its population; Karachi is a megacity with around 20 million inhabitants. It is the trading and financial hub of the country, and receives more local and international travelers than other regions. Further, higher case reports may be attributed to the extensive network of diagnostic laboratories in Karachi. Notably, the case numbers reported between July and August 2022 surge were lower than those reported between January and February 2022. Possibly, due to reduced testing rates within the population. Limited testing during the latter period could be attributed to reduced disease severity of COVID-19 from Omicron variants and increased vaccination coverage and thus less concern in the population about symptomatic infections [28].

There is a dearth of information regarding SARS-CoV-2 genomic surveillance from Asia, with studies providing insights into viral transmission from limited datasets [29].

It is a limitation of this study that interpretations of the impact of Omicron subvariants in the different provinces of Pakistan was dependent on genome submissions from each region, and that these were not consistent. For instance, despite having identified Sindh as the entry-point of viral strains, this could be skewed by the limited data from other provinces. Another consequence of the limited genomic surveillance is delayed sampling, testing, and reporting. We used the sample collection dates for phylodynamic analysis and are able to provide insights regarding strain variations across the study period. Additionally, the sequencing was not consistent over the entire study period, as more samples were sequenced in the wave between December 2021 and February 2022. This makes it difficult to analyze the data in the context of burden of disease with age and gender stratification. However, given that most of the COVID-19 cases reported from Pakistan are from those aged 40 years and below, it is not surprising that we found that there was a greater representation of Omicron variants in this age group. This is in keeping with the younger age of the population of Pakistan, with 65% of individuals aged below 30 years. Another limitation of the study is that only the first Omicron subvariant, BA.1, was studied in terms of phylogenetic relatedness. Overall, it is likely that BA.5 samples had a greater representation in this study selection as the increase in this VoC occurred during a period in 2022 when there had been an increase in sequencing

capacity for SARS-CoV-2 genomics. In the context of Aga Khan University, this was due to a contribution of institutional (Aga Khan University), national (GCF grant no.913, Higher Education Commission, Pakistan; World Health Organization, Pakistan) and international (Health Security Partners, USA; Fogarty International Center, NIH, USA; Bill and Melinda Gates Foundation) funding support for sequencing and bioinformatics initiatives.

In conclusion, correlation of SARS-CoV-2 genomic data with COVID-19 epidemiological data in Pakistan allowed us to describe the introduction of Omicron subvariants with different pandemic waves. Further, information regarding the relatedness of the lineages introduced in each province provides insights into the possible reasons for the waves observed in the provinces, which were led by Sindh and then followed in each case by ICT and Punjab. There is a need to establish more robust genomic surveillance networks to adequately represent the entire country. The importance of continuing genomic surveillance matched with analyzing epidemiological data is essential for successful management of a highly transmissible pathogen such as SARS-CoV-2.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09539-3>.

Additional file 1.

Additional file 2: Supplementary Figure 1. SARS-CoV-2 genome submissions in GISAID from Pakistan.

Additional file 3: Supplementary Figure 2. SARS-CoV-2 genome submissions in GISAID from each region of Pakistan across time.

Acknowledgements

Not applicable.

Authors' contributions

ARB, YAR, JA and ZH planned the study. ARB, AK, JA, PMT, BM, NST, WUK were involved in sequencing and bioinformatics analysis. YAR, AK, MY and ARB conducted the data analysis. ZH, RH, IN, WUK, ZR, DS and UBA got funding support for this work. ZR, NST and DS edited the paper. All authors approved the manuscript.

Funding

Research support was provided by the Aga Khan University, Pakistan; World Health Organization, Islamabad, Pakistan; Higher Education Commission, Pakistan; Health Security Partners, USA; and a Bill and Melinda Gates Foundation grant. Fogarty International Center, National Institutes of Health, USA provided training for genomic epidemiology. 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.

Availability of data and materials

The detailed raw data will be available on request to the corresponding author.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethical Review Committee, The Aga Khan University (AKU). No need for consent as the data from online resource GISAID is used for the analysis.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Pathology and Laboratory Medicine, Aga Khan University, Karachi 74800, Pakistan. ²Fogarty International Center, U.S. National Institutes of Health, 16 Center Drive, Bethesda, MD 20892, USA. ³Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA. ⁴Department of Pediatrics and Child Health, Aga Khan University, Karachi 74800, Pakistan. ⁵World Health Organization Country Office, Park Road, Chak Shahzad, Islamabad, Pakistan.

Received: 30 April 2023 Accepted: 27 July 2023

Published online: 03 August 2023

References

- Gonzalez-Candelas F, et al. One year into the pandemic: Short-term evolution of SARS-CoV-2 and emergence of new lineages. *Infect Genet Evol.* 2021;92: 104869.
- Go. COVID-19 Health Advisory Platform. Ministry of National Health Services Regulations and Coordination. 2021 December 2022]; Available from: <https://covid.gov.pk/>.
- JHU. 2020; Available from: <https://coronavirus.jhu.edu/region/pakistan>. Cited 2022.
- Chen X, et al. Impact of vaccination on the COVID-19 pandemic in US states. *Sci Rep.* 2022;12(1):1–10.
- Tan ST, et al. COVID-19 Vaccination and Estimated Public Health Impact in California. *JAMA Netw Open.* 2022;5(4):e228526–e228526.
- Bell E, et al. Estimates of the Global Burden of COVID-19 and the Value of Broad and Equitable Access to COVID-19 Vaccines. *Vaccines (Basel).* 2022;10(8):1320.
- Yang Z, et al. Clinical characteristics, transmissibility, pathogenicity, susceptible populations, and re-infectivity of prominent COVID-19 variants. *Aging Dis.* 2022;13(2):402–22.
- Khandia R, et al. Emergence of SARS-CoV-2 Omicron (B.1.1.529) variant, salient features, high global health concerns and strategies to counter it amid ongoing COVID-19 pandemic. *Environ Res.* 2022;209:112816.
- Batra K, et al. Evolution of SARS-CoV-2 variants: a rapid literature scan. *J Health Soc Sci.* 2022;7(2):141–51.
- Zahmatkesh S, et al. Review of concerned SARS-CoV-2 variants like Alpha (B. 1.1. 7), Beta (B. 1.351), Gamma (P. 1), Delta (B. 1.617. 2), and Omicron (B. 1.1. 529), as well as novel methods for reducing and inactivating SARS-CoV-2 mutants in wastewater treatment facilities. *J Haz Mat Adv.* 2022:100140.
- WHO. One year since the emergence of COVID-19 virus variant Omicron. 2022.
- Khan S, et al. The Burden of Omicron Variant in Pakistan: An Updated Review. *COVID.* 2022;2(10):1460–76.
- Imran M, et al. COVID-19 situation in Pakistan: a broad overview. *Respirology (Carlton, Vic).* 2021;26:891.
- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* 2020;20(5):533–4.
- Nasir A, et al. Tracking SARS-CoV-2 variants through pandemic waves using RT-PCR testing in low-resource settings. *PLOS Glob Public Health.* 2023;3(6):e0001896.
- Umair M, et al. Tracking down B.1.351 SARS-CoV-2 variant in Pakistan through genomic surveillance. *J Med Virol.* 2022;94(1):32–4.

17. Huddleston J, et al. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *J Open Source Softw.* 2021;6(57):2906.
18. Katoh K, et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059–66.
19. Minh BQ, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37(5):1530–4.
20. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus evolution.* 2018;4(1):vex042.
21. Hadfield J, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics.* 2018;34(23):4121–3.
22. JHU. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University: John Hopkins University. 2021; Available from: <https://github.com/CSSEGISandData/COVID-19>.
23. Jamil B, et al. Interferon γ :L10 ratio defines the severity of disease in pulmonary and extra-pulmonary tuberculosis. *Tuberculosis.* 2007;87:279–87.
24. Nasir N, et al. Clinical characteristics and outcomes of COVID-19: Experience at a major tertiary care center in Pakistan. *J Infect Dev Ctries.* 2021;15(4):480–9.
25. Asif N, Younas A, Gilani M, Shaikh W, Ain QU. Experience Of Severe Acute Respiratory Syndrome Coronavirus-2 (Sars-Cov-2) - Covid-19 at a Tertiary Care Hospital in Quetta Baluchistan. *Pak Armed Forces Med J.* 2021;71:2152.
26. Peshawar records highest mortality rate. *DAWN* 2020 April 27, 2020; Available from: <https://www.dawn.com/news/1552545>.
27. Nasir A, et al. Evolutionary history and introduction of SARS-CoV-2 Alpha VOC/B. 1.1. 7 in Pakistan through international travelers. *Virus Evol.* 2022;8(1):veac020.
28. Nisar MI, et al. Assessing the effectiveness of COVID-19 vaccines in Pakistan: a test-negative case-control study. *J Infect.* 2023;86:e144.
29. Zhu M, et al. Molecular phylogenesis and spatiotemporal spread of SARS-CoV-2 in Southeast Asia. *Front Public Health.* 2021;9:685315.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

