

1 **Sequential viral introductions and spread of BA.1 drove the Omicron wave across**
2 **Pakistani provinces**

3 Ali Raza Bukhari¹, Javaria Ashraf¹, Akbar Kanji¹, Yusra Abdul Rahman¹, Nídia S. Trovãõ³,
4 Peter M. Thielen², Maliha Yameen¹, Samiah Kanwar⁴, Waqasuddin Khan⁴, Furqan Kabir⁴, M.
5 Imran Nisar⁴, Brian Merritt², Rumina Hasan¹, David Spiro³, Zeba Rasmussen³, Uzma Bashir
6 Aamir⁵, Zahra Hasan^{1#}

7 ¹Department of Pathology and Laboratory Medicine, Aga Khan University, Karachi-74800,
8 Pakistan

9 ²Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD
10 20723, USA

11 ³Fogarty International Center, U.S. National Institutes of Health, 16 Center Drive, Bethesda, MD
12 20892, USA

13 ⁴Department of Pediatrics and Child Health, Medical College, Aga Khan University, Karachi-
14 74800, Pakistan

15 ⁵World Health Organization country office, Park Road, Chak Shahzad, Islamabad, Pakistan

16 **#Corresponding author**

17 Zahra Hasan, PhD, Professor, Department of Pathology and Laboratory Medicine, Aga Khan
18 University, Stadium Road, P.O.Box 3500, Karachi 74800, Pakistan. Email:
19 zahra.hasan@aku.edu

20

21 **Abstract**

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

25 Methods: We studied 276,525 COVID-19 cases and 1,041 genomes sequenced from December
26 2021 to August 2022. Sequences were analyzed and visualized using phylogenetic trees.

27 Results: The highest case numbers and deaths were recorded in Sindh and Punjab, the most
28 populous provinces in Pakistan. Omicron variants comprised 95% of all genomes, with BA.2
29 (34.2%) and BA.5 (44.6%) predominating. The first Omicron wave was associated with the
30 sequential identification of BA.1 in Sindh, then Islamabad Capital Territory, Punjab, Khyber
31 Pakhtunkhwa (KP), Azad Jammu Kashmir (AJK), Gilgit-Baltistan (GB) and Balochistan.
32 Phylogenetic analysis revealed Sindh to be the source of BA.1 and BA.2 introductions into
33 Punjab and Balochistan during early 2022. BA.4 was first introduced in AJK and BA.5 in
34 Punjab. Most recent common ancestor (MRCA) analysis revealed relatedness between the
35 earliest BA.1 genome from Sindh with Balochistan, AJK, Punjab and ICT, and that of first BA.1
36 from Punjab with strains from KPK and GB.

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

41 **Keywords:** Omicron; variants of concern; SARS-CoV-2; phylogenetics; transmission dynamics

42 **Introduction**

43 The rise and fall of coronavirus disease 2019 (COVID-19) cases since the emergence of severe
44 acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in December 2019 has driven the need
45 for monitoring of variants through genomic surveillance [1]. The global burden of COVID-19 is
46 not fully known although greater than 754 million cases have been reported as of February 3,
47 2022. Pakistan has reported about 1.5 million cases of COVID-19 and nearly 31,000 deaths [2,
48 3]. COVID-19 vaccines have had a significant impact on controlling both morbidity and
49 mortality from COVID-19 [4, 5]. The proportion of deaths to cases changed greatly throughout
50 the pandemic with reduced disease severity after vaccines were introduced and the evolution of
51 SARS-CoV-2 variants of concern (VoC) toward higher transmissibility and lower pathogenicity.
52 The epidemiology of COVID-19 has been informed by the reported number of cases, deaths,
53 hospitalizations, and viral genomic sequencing. This information has varied greatly between
54 high- and low- income countries of similar population sizes; explanatory factors include lack of
55 resources needed for gathering metadata, PCR testing and genomic sequencing [6].

56 SARS-CoV-2 VoCs have shown a trend toward increased transmissibility, leading to strain-
57 specific increases in COVID-19 cases [7]. The first VoC was the Alpha variant or B.1.1.7,
58 followed by Beta/B.1.351 and Delta/B.1.617.2 in 2021. VoCs which emerged after the
59 introduction of vaccinations were more effective at evading host immunity driven by both
60 natural infection and vaccinations [8]. This was evidenced by Omicron and its subvariants
61 coming to dominate over other lineages across the globe by the end of 2021 [9, 10]. Omicron has
62 distinct subvariants, five of which include BA.1, BA.2, BA.3, BA.4 and BA.5 [11]. BA.1 was
63 first identified in South Africa and Botswana on November 26, 2021 [12]. By January 2022,
64 BA.1 made up 81% of cases in South Africa, rapidly decreasing to 45% in February 2022. BA.5
65 was first identified in South Africa on January 21, 2022, and spread elsewhere, particularly in
66 Europe, by April 2022. Soon after this time, BA.4 and BA.5 variants became predominant [13].

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

72 health officials can make decisions about the allocation of resources and implementation of
73 preventive measures to limit the spread of the virus.

74 Pakistan has a population of approximately 220 million people. It consists of seven distinct
75 regional territories: Azad Jammu and Kashmir (AJK), Balochistan, Gilgit-Baltistan (GB),
76 Khyber Pakhtunkhwa (KP), Punjab, Sindh and the Islamabad Capital Territory (ICT). The first
77 COVID-19 case from Pakistan was reported in Sindh on February 26, 2020. The timeline of
78 pandemic waves experienced in the country was; March to July 2020 where the G Nextstrain
79 clade of SARS-CoV-2 dominated, October 2020 to January 2021 dominated by GR/GH clades,
80 April to May 2021 where Alpha dominated [14], July to September 2021 where Delta was the
81 predominant strain, and, December 2021 to February 2022 dominated by Omicron.

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

85 **Materials and Methods**

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

87 **Selection of SARS-CoV-2 genomes**

88 This was an analysis of SARS-CoV-2 genomes submitted of Pakistani origin. Sequences with
89 collection dates between December 1, 2021 and August 14, 2022 were obtained from GISAID on
90 September 15, 2022 “complete” and “low coverage excluded” filtering criteria. We downloaded
91 1041 sequences with metadata; of these, 984 were found to belong to the Omicron VoC, of
92 which 957 had complete metadata and were used for phylogenetic analysis.

93 **COVID-19 cases and mortality data**

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

98 **Phylogenetic tree and analysis**

99 FASTA files of the 984 Omicron sequences with corresponding metadata (age, gender, date of
100 collection and location) were used for phylogenetic tree reconstruction using the augur pipeline
101 [15]. Out of the total (984), 957 genomic sequences qualified for phylodynamic mapping.

102 Full length SARS-CoV-2 genomes of Omicron subvariants were aligned using the MAFFT
103 alignment tool [16]. Multiple Sequence Alignment (MSA) files generated from MAFFT were
104 used for a maximum likelihood (ML) phylogenetic tree through IQ-TREE2 [17]. By applying a
105 generalized midpoint rooting strategy, rooting of the tree was carried out with branch length
106 variance using TreeTime [18]. The tree was visualized and edited in Figtree v. 1.4.4
107 (<http://tree.bio.ed.ac.uk>). The final tree with annotated nodes and metadata was exported to the
108 phylodynamic visualizing tool Auspice [19].

109 **Statistical Analysis**

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

113 **Results**

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

115 A total of 276,525 cases were reported between December 2021 and August 2022 [20]. The
116 region-wide distribution of COVID-19 cases is depicted in Figure 1, which also depicts the
117 population density of each of the 7 regions studied. Sindh reported the most cases (41.7%),
118 followed by Punjab with 27.1%, 15.2% from KP, 11% from ICT, 3.4% from AJK, 0.9% from
119 Balochistan and 0.6% from GB of total cases.

120 A total of 1,766 COVID-19 deaths were reported during the study period (Figure 1). Sindh
121 reported the greatest (33.1% of total deaths) followed by Punjab (31.1%), KP (26.7%), ICT
122 (3.9%), AJK (2.9%), Balochistan (2.0%) and GB (0.3%). The overall case fatality ratio
123 percentage (CFR%) over this period was 0.6%. It was highest in Balochistan (1.5%), followed by
124 KP (1.1%), AJK (0.5%), Sindh (0.5%), Punjab (0.5%), GB (0.3%) and ICT (0.2%).

125 COVID-19 case numbers rose rapidly from under 1,000 to more than 10,000 per week across
126 Pakistani regions between the end of January and beginning of February 2022, with 22,000 cases

127 being reported in the last week of January 2022 in Sindh alone (Figure 2). COVID-19 peaks
128 occurred sequentially in other provinces; ICT by the last week of January 2022; Punjab, KP,
129 Balochistan and AJK by the first week of February 2022; and GB by the second week of
130 February 2022. Cases declined around early March 2022, reaching a few hundred cases per week
131 by June 2022. Another slight rise in cases was observed in Sindh (early July 2022) and Punjab
132 (early August 2022). A similar trend was also observed in other regions. Of note, a rise in
133 COVID-19 cases was observed in Sindh ahead of the other regions of Pakistan.

134 The mean age of COVID-19 cases across Pakistan was 39 SD19 years. For each region the mean
135 age was: ICT 38 SD19, Punjab 39 SD21, Sindh 40 SD19, KP 32 SD19, Balochistan 48 SD17,
136 AJK 38 SD19 and GB 43 SD24 years. There was no significant difference between age groups
137 of COVID-19 cases across the regions.

138 **Heterogeneity in data submission across regions of Pakistan**

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

148 The pango lineage distribution was A (0.3%), AY (4.7%), B (1.7%), B.1 (0.8%), BA.1
149 (Nextstrain 21K; 16.6%), BA.2 (Nextstrain clade 21L; 32.2%), BA.4 (Nextstrain 22A; 3.6%) and
150 BA.5 (Nextstrain 22B; 40.3%).

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

155 **Omicron and subvariants across regions of Pakistan**

156 Phylogenetic analysis of Omicron subvariant sequences is depicted in Figure 3. ICT uploaded
157 363 sequences to GISAID, the greatest of which were BA.5 (n=161, 44.4%) and BA.2 (n=153,
158 42.1%). There were 139 sequences from Punjab, predominantly BA.5 (n=58, 41.7%) and BA.2
159 (n=52, 37.4%). The distribution of 282 genomes from Sindh included BA.5 (n=131, 46.5%),
160 BA.2 (n=63, 22%) and BA.1 (n=76, 27%). The fewest genomes were submitted from AJK
161 (n=63, 6%), KP (n=61, 6%) and GB (n=43, 4%); of those available, BA.2 and BA.5 were the
162 predominant subvariants. The fewest sequences were from Balochistan (n=14, 2%), which did
163 not report any BA.5 subvariant.

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

165 The phylogenetic analysis of the Omicron genomes across Pakistan is depicted in Figure 4. The
166 phylogram presents the evolution and spread over time of BA.1, BA.2, BA.4 and BA.5 variants
167 across the country, including sequences from ICT, Sindh, Punjab, AJK, KP, GB and Balochistan.
168 Some were from travelers who entered Sindh from the Kingdom of Saudi Arabia (KSA, n=3),
169 the United Arab Emirates (UAE, n =2), India (n=1), Turkey (n=1) and the United States of
170 America (USA, n=1).

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

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

175 The first case of BA.1 was identified in Sindh (Figure 5A). The BA.1 lineage encompassed sub-
176 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
177 BA.1.18. Subsequently, BA.1 was reported in Punjab, ICT, and AJK. Later, reports of BA.1
178 were obtained from KP (January 2022) and Balochistan (February 2022).

179 We next examined the phylogeny of the 333 BA.2 sequences found between January and July
180 2022. The first BA.2 subvariant was reported in January 2022 from Sindh, and clustered closely
181 with an isolate identified in a traveler from India (Figure 5B). Subsequently, BA.2 was reported
182 in ICT and Punjab in April 2022.

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

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

187 **Phylogenetic relatedness of cases reported across different regions of Pakistan**

188 To further understand the genetic relatedness of Omicron subvariants across Pakistan, we
189 analyzed the relatedness of sequenced genomes through the inference of the most recent
190 common ancestor (MRCA) across all 957 genomes (Figure 6A), focusing on the first BA.1 case
191 reported in each region. The first BA.1 case from Balochistan, with a sampling date of January
192 21, 2022, shared a common ancestral node with one reported earlier from Sindh on January 10,
193 2022 (Figure 6B). The first case from Sindh, reported on December 8, 2022, had no known
194 association with an overseas traveler [21] (Figure 6C). The first BA.1 case from AJK (January 8,
195 2022) shared an ancestral node with a case from Sindh and Punjab (Figure 6D). The earliest case
196 from GB and KP had the same MRCA and shared an ancestral node with isolates from the
197 Punjab (Figure 6E). Similarly, the earliest BA.1 cases from ICT and Punjab shared an ancestral
198 node with Sindh (Figure 6F).

199 **Discussion**

200 Genomic surveillance data from Pakistan have been limited, especially for the earlier pandemic
201 waves of 2020 and 2021. Our study provides insights into the phylogenetic relatedness of
202 different Omicron variants which spread across Pakistan. Here, we associate the introduction of
203 the Omicron subvariants with COVID-19 surges in the different regions across the country. This
204 is the first study highlighting the genetic relatedness of Omicron subvariants in Pakistan through
205 an analysis of the MRCAs. We identify Sindh as a hotspot for variant introductions into the
206 country.

207 COVID-19 waves were found to display a pattern of sequential geographic transmission,
208 occurring first in Sindh, followed by Punjab and other regions. Overall, the majority of COVID-
209 19 cases were aged below 40 years. The predominant age group of cases was 19-40 years.
210 Deaths have previously been associated with older age groups [22]. During the period covered
211 by this study, the CFR was found to be the highest in Balochistan (1.5%), 1.1% in KP and with

212 all other regions reporting a CFR of less than 1%. Over the course of the COVID-19 pandemic,
213 the highest mortality (3.5%) reported in Pakistan was from Peshawar, KP, in 2020. Local
214 experts have suggested the lack of social distancing and a non-compliant response to standard
215 operating procedures (SOPs) in the community to be the major reason behind this [23]. Other
216 factors potentially contributing to the higher CFR could include a delayed presentation of
217 COVID-19 symptoms combined with limited testing and reporting.

218 Balochistan reported the fewest cases, and the fewest genomic sequences were submitted from
219 the region. Hence, due to the limited data available it is not possible to understand fully the trend
220 of COVID-19 in Balochistan. One study showed COVID-19 positivity of 13% between March
221 and December 2020, with the mean age of positive cases to be 36 ± 14 years, with 20% of the
222 cases being female [24]. The lack of a provincial sequencing facility in Balochistan limits the
223 available sequencing efforts. Another challenge regarding COVID-19 case information in
224 Balochistan is the limited access to healthcare facilities for this large but sparsely populated
225 province (6% of population with 43.6% of the land area of Pakistan).

226 Omicron variants comprised most of the sequences submitted to GISAID during the study
227 period. BA.5 and BA.2 variants were the most prevalent overall amongst the genomes analyzed.
228 We tried to determine if there was an association between the introduction of the first Omicron
229 subvariant and the increase in number of cases across geographical regions of Pakistan. The first
230 introduction was BA.1 on 8th December 2021 into Karachi, Sindh. BA.1 was reported later from
231 Punjab, ICT and AJK. BA.1 reports from Balochistan and KP occurred after January 2022. Soon
232 after identification of BA.1 strains, Pakistan experienced a surge of cases across the country,
233 with particularly high numbers of cases and associated deaths in Karachi, Sindh. The
234 identification of a larger than average case count in the Sindh region was associated with
235 travelers from other destinations. Earlier, the introduction of the Alpha VoC was associated with
236 international travelers to Karachi [25]. The high case count in Sindh could be also driven by the
237 large population size; Karachi is a megacity with around 20 million inhabitants and, as the
238 trading and financial hub of the country, receives more local and international travelers than
239 other regions.

240 The first case from Sindh appeared to have transmitted locally as reported in previous studies
241 [21]. The earliest BA.1 cases from ICT and Punjab share an ancestral node with Sindh, which

242 does not necessarily mean that the variant spread through travel between these regions, but could
243 also be attributed to an independent introduction from elsewhere.

244 Phylogenetic analysis showed that the first BA.1 strain from Balochistan was related to that from
245 Sindh. The provinces have a shared border with frequent routine travel between them, allowing
246 easy spread, however phylogeographic analysis would be needed to test this hypothesis.

247 The first BA.1 case from AJK (January 8, 2022) shared an ancestral node with a case from Sindh
248 and Punjab. The earliest case from GB and KP had the same MRCA and shared an ancestral
249 node with isolates from the Punjab. BA.1 cases in GB, KP, and AJK regions shared a same
250 ancestral node, which might be due to the proximity of these three provinces in the northern
251 region of Pakistan.

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

257 It is likely that the rise in cases observed in Sindh (July 2022), Punjab (August 2022) and similar
258 trends observed in other regions was due to the spread of BA.4 and BA.5 strains at this time.
259 Notably, the increase in cases was first reported in Sindh. This is likely due to both its population
260 size (mainly Karachi) and the extensive network of diagnostic laboratories in the region. The
261 case numbers of the July–August 2022 surge were less than those of the January–February 2022
262 surge, possibly due to reduced testing rates within the population. Limited testing during the
263 latter period could be attributed to reduced disease severity of COVID-19 from Omicron variants
264 and increased vaccination coverage and thus less concern in the population about symptomatic
265 infections.

266 There is a dearth of information regarding SARS-CoV-2 genomic surveillance from Asia, with
267 studies providing insights into viral transmission from limited datasets[26]. This is the first study
268 to conduct a detailed phylogenetic analysis of Omicron subvariants in Pakistan to understand
269 their first introductions. Sindh had the highest number of reported COVID-19 cases (41.7% of

270 total cases), while its population size is less than half of Punjab's, suggesting that the rate of
271 COVID-19 cases on a population basis was higher in Sindh than in Punjab.

272 Similarly, Balochistan had the lowest number of reported COVID-19 cases (0.9% of total cases),
273 but its population size is also the smallest among all regions. However, the CFR in Balochistan
274 was the highest (1.5%) among all regions, indicating that the pandemic's impact was severe
275 there. In a similar manner, KP has a population of 36 million but has much higher deaths as
276 compared to Sindh (48 million) and Punjab (110 million), as reflected by the CFR (1.1%)

277 The availability of sequencing data varies widely between different regions in Pakistan. It is a
278 limitation of this study that full interpretations of the impact of Omicron subvariants in the
279 different provinces of Pakistan cannot be made due to limited genomic surveillance in some
280 regions. For instance, despite having identified Sindh as the entry-point of viral strains, this
281 could be skewed by the limited data from other provinces. Another consequence of the limited
282 genomic surveillance is delayed sampling, testing, and reporting. We used the sample collection
283 dates for phylodynamic analysis and are able to provide insights regarding strain variations
284 across the study period. Additionally, the sequencing was not consistent over the entire study
285 period, as more samples were sequenced in the wave between December 2021 and February
286 2022. This makes it difficult to analyze the data in the context of burden of disease with age and
287 gender stratification. However, given that most of the COVID-19 cases reported from Pakistan
288 are from those aged 40 years and below, it is not surprising that we found that there was a greater
289 representation of Omicron variants in this age group. This is in keeping with the younger age of
290 the population of Pakistan, with 65% of individuals aged below 30 years. Another limitation of
291 the study is that only the first Omicron subvariant, BA.1, was studied in terms of phylogenetic
292 relatedness. Overall, it is likely that BA.5 samples had a greater representation in this study
293 selection as the increase in this VoC occurred during a period in 2022 when there had been an
294 increase in sequencing capacity for SARS-CoV-2 genomics. In the context of Aga Khan
295 University, this was due to a contribution of institutional (Aga Khan University), national
296 (Higher Education Commission, Pakistan; World Health Organization, Pakistan) and
297 international (Health Security Partners, USA; Fogarty International Center, NIH, USA; Bill and
298 Melinda Gates Foundation) funding support for sequencing and bioinformatics initiatives.

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

308 **Acknowledgements**

309 Research support was provided by the Aga Khan University, Pakistan; World Health
310 Organization, Islamabad, Pakistan; Higher Education Commission, Pakistan; Health Security
311 Partners, USA; and a Bill and Melinda Gates Foundation grant. Fogarty International Center,
312 National Institutes of Health, USA provided training for genomic epidemiology. The opinions
313 expressed in this article are those of the authors and do not reflect the view of the National
314 Institutes of Health, the Department of Health and Human Services, or the United States
315 government.

316 **Authors contributions**

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

321 References

- 322 1. Gonzalez-Candelas, F., et al., *One year into the pandemic: Short-term evolution of SARS-CoV-2*
323 *and emergence of new lineages*. *Infect Genet Evol*, 2021. **92**: p. 104869.
- 324 2. *Go. COVID-19 Health Advisory Platform. Ministry of National Health Services Regulations and*
325 *Coordination*. 2021 December 2022]; Available from: <https://covid.gov.pk/>.
- 326 3. JHU. 2020 [cited 2022; Available from: <https://coronavirus.jhu.edu/region/pakistan>.
- 327 4. Chen, X., et al., *Impact of vaccination on the COVID-19 pandemic in US states*. *Scientific reports*,
328 2022. **12**(1): p. 1-10.
- 329 5. Tan, S.T., et al., *COVID-19 Vaccination and Estimated Public Health Impact in California*. *JAMA*
330 *network open*, 2022. **5**(4): p. e228526-e228526.
- 331 6. Bell, E., et al., *Estimates of the Global Burden of COVID-19 and the Value of Broad and Equitable*
332 *Access to COVID-19 Vaccines*. *Vaccines (Basel)*, 2022. **10**(8).
- 333 7. Yang, Z., et al., *Clinical Characteristics, Transmissibility, Pathogenicity, Susceptible Populations,*
334 *and Re-infectivity of Prominent COVID-19 Variants*. *Aging Dis*, 2022. **13**(2): p. 402-422.
- 335 8. Khandia, R., et al., *Emergence of SARS-CoV-2 Omicron (B.1.1.529) variant, salient features, high*
336 *global health concerns and strategies to counter it amid ongoing COVID-19 pandemic*. *Environ*
337 *Res*, 2022. **209**: p. 112816.
- 338 9. Batra, K., et al., *Evolution of SARSCoV-2 variants: A rapid literature scan*. *J Health Soc Sci*, 2022.
339 **7**(2): p. 141-151.
- 340 10. Zahmatkesh, S., et al., *Review of concerned SARS-CoV-2 variants like Alpha (B.1.1.7), Beta*
341 *(B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529), as well as novel methods for*
342 *reducing and inactivating SARS-CoV-2 mutants in wastewater treatment facilities*. *Journal of*
343 *Hazardous Materials Advances*, 2022(2772-4166 (Print)).
- 344 11. Nasir, A., et al., *SARS-CoV-2 Variants of Concern (VOC) Alpha, Beta, Gamma, Delta, and Omicron*
345 *coincident with consecutive pandemic waves in Pakistan*. *medRxiv*, 2022.
- 346 12. WHO, *One year since the emergence of COVID-19 virus variant Omicron*. 2022.
- 347 13. Khan, S., et al., *The Burden of Omicron Variant in Pakistan: An Updated Review*. *COVID*, 2022.
348 **2**(10): p. 1460-1476.
- 349 14. Imran, M., et al., *COVID-19 situation in Pakistan: A broad overview*. *Respirology (Carlton, Vic.)*,
350 2021.
- 351 15. Huddleston, J., et al., *Augur: a bioinformatics toolkit for phylogenetic analyses of human*
352 *pathogens*. *J Open Source Softw*, 2021. **6**(57).
- 353 16. Katoh, K., et al., *MAFFT: a novel method for rapid multiple sequence alignment based on fast*
354 *Fourier transform*. *Nucleic acids research*, 2002. **30**(14): p. 3059-3066.
- 355 17. Minh, B.Q., et al., *IQ-TREE 2: new models and efficient methods for phylogenetic inference in the*
356 *genomic era*. *Molecular biology and evolution*, 2020. **37**(5): p. 1530-1534.
- 357 18. Sagulenko, P., V. Puller, and R.A. Neher, *TreeTime: Maximum-likelihood phylodynamic analysis*.
358 *Virus evolution*, 2018. **4**(1): p. vex042.
- 359 19. Hadfield, J., et al., *Nextstrain: real-time tracking of pathogen evolution*. *Bioinformatics*, 2018.
360 **34**(23): p. 4121-4123.
- 361 20. JHU. *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at*
362 *Johns Hopkins University: Johns Hopkins University*. 2021; Available from:
363 <https://github.com/CSSEGISandData/COVID-19>.
- 364 21. *Pakistan confirms first case of Omicron variant*. Available from:
365 [https://www.reuters.com/world/asia-pacific/pakistan-confirms-first-case-omicron-variant-2021-](https://www.reuters.com/world/asia-pacific/pakistan-confirms-first-case-omicron-variant-2021-12-13/)
366 [12-13/](https://www.reuters.com/world/asia-pacific/pakistan-confirms-first-case-omicron-variant-2021-12-13/).

- 367 22. Nasir, N., et al., *Clinical characteristics and outcomes of COVID-19: Experience at a major tertiary*
368 *care center in Pakistan*. J Infect Dev Ctries, 2021. **15**(4): p. 480-489.
- 369 23. *Peshawar records highest mortality rate*. DAWN 2020 April 27, 2020; Available from:
370 <https://www.dawn.com/news/1552545>.
- 371 24. Naveed Asif, S.A.S., Alveena Younas, Mehwish Gilani, Waqas Shaikh, Qurat Ul Ain, *Experience Of*
372 *Severe Acute Respiratory Syndrome Coronavirus-2 (Sars-Cov-2) - Covid-19 at a Tertiary Care*
373 *Hospital in Quetta, Baluchistan*. Pakistan Armed Forces Medical Journal, 2021. **Vol 71**
- 374 25. Nasir, A., et al., *Evolutionary history and introduction of SARS-CoV-2 Alpha VOC/B. 1.1. 7 in*
375 *Pakistan through international travelers*. Virus evolution, 2022. **8**(1): p. veac020.
- 376 26. Zhu, M., et al., *Molecular Phylogenesis and Spatiotemporal Spread of SARS-CoV-2 in Southeast*
377 *Asia*. Front Public Health, 2021. **9**: p. 685315.
- 378

379 **Figure Legends**

380 **Figure 1. Population density, COVID -19 cases, deaths and CFR% across Pakistan.** The
381 graph depicts nationwide data from December 1, 2021, until August 14, 2022. A. The region-
382 wise population density of Azad Jammu and Kashmir (AJK), Balochistan, Gilgit-Baltistan (GB),
383 Islamabad Capital Territory (ICT), Khyber Pakhtunkhwa (KPK), Punjab and Sindh is presented.
384 The scale bar displays population values in millions of persons shaded by color. B. Left panel x-
385 axis shows number of COVID-19 cases, middle panel shows number of COVID-19 related
386 deaths and right panel shows CFR% of regions represented by color: ICT (light blue), Punjab
387 (dark blue), Sindh (green), KPK (orange), Balochistan (purple), GB (yellow) and AJK (red).

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

392 **Figure 3. Frequency of Omicron variants across Pakistan.** The graph shows all Omicron
393 (n=984) genomes submitted from each region between December 1, 2021, and August 14, 2022.
394 BA.1 (blue), BA.2 (orange), BA.4 (yellow) and BA.5 (green).

395 **Figure 4. Phylogenetics of early Omicron variants in Pakistan.** The tree illustrates the
396 relatedness of 957 omicron sequences submitted from Pakistan between December 1, 2021, and
397 August 14, 2022. The tree uses color coding to identify the travel origin of each case reported.

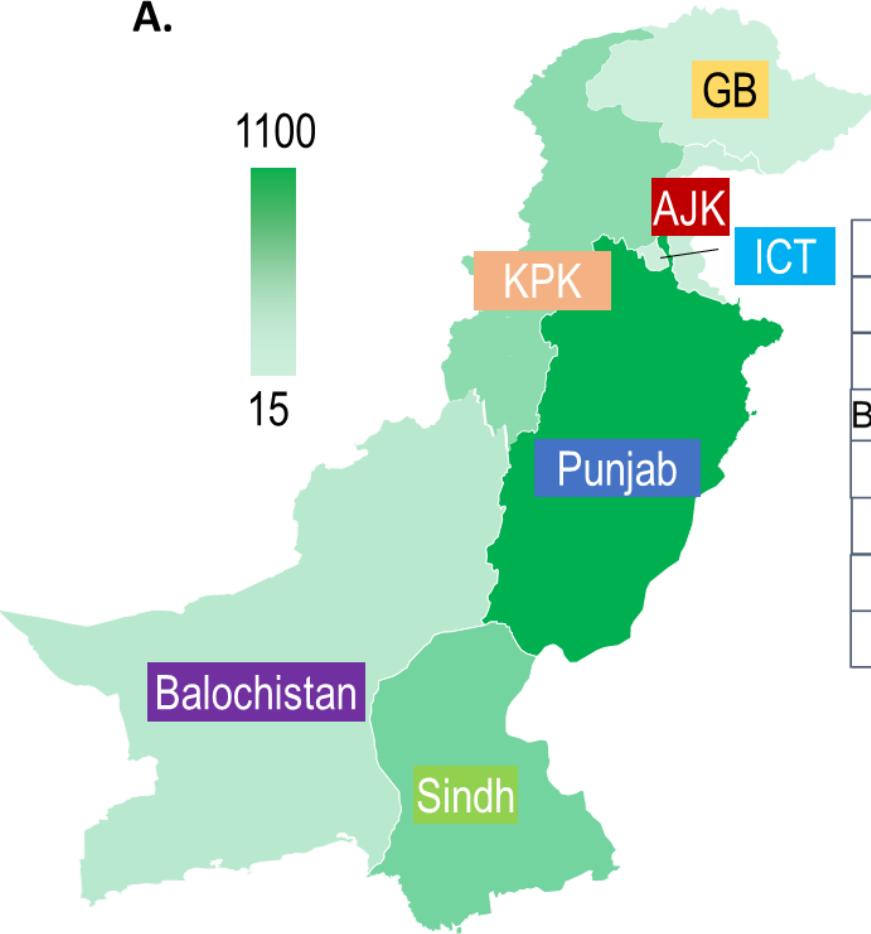
398 **Figure 5. Introduction and linkage of omicron variants in Pakistan.** Phylogenetic trees depict
399 the first case report for each variant as a red circle in the identified timeline with relatedness to
400 later isolates. Panels A-D depict the first case of each variant identified by a red circle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

404 **Figure 6. Evolutionary relatedness of the first reported Omicron variant from each region**
405 **of Pakistan.** A. Phylogenetic tree for 957 Omicron genomes and related strains sharing the most
406 recent common ancestor (MRCA). The first BA.1 sequence from each region, B, Balochistan
407 (purple); C, Sindh (green); D, AJK (light red); E, Gilgit-Baltistan and KPK (yellow); F, Punjab

408 and ICT (light blue). Data are presented as Figtree v. 1.4.4 outputs using the maximum-
409 likelihood tree generated using IQTREE v. 2.2.0. Grey triangles represent collapsed nodes of
410 relatively distant sequences from the first Omicron sequences in each region. Scale bar presents
411 nucleotide substitution/site.

Figure 1

A.



B.

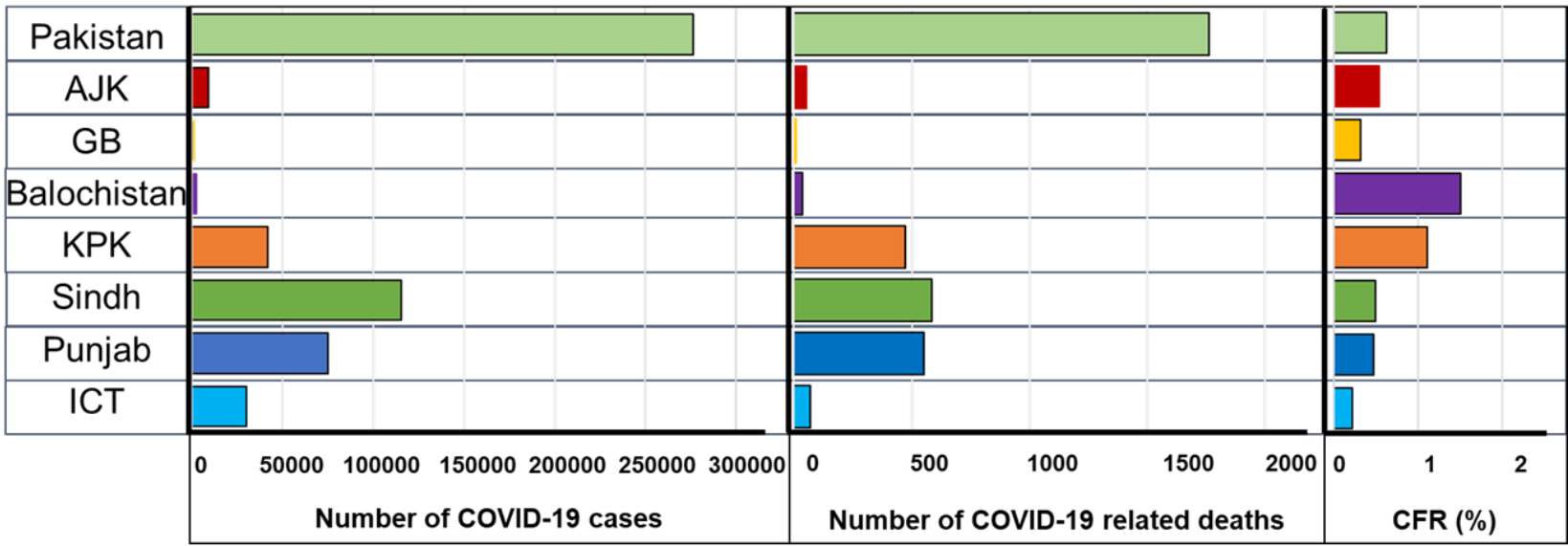


Figure 2

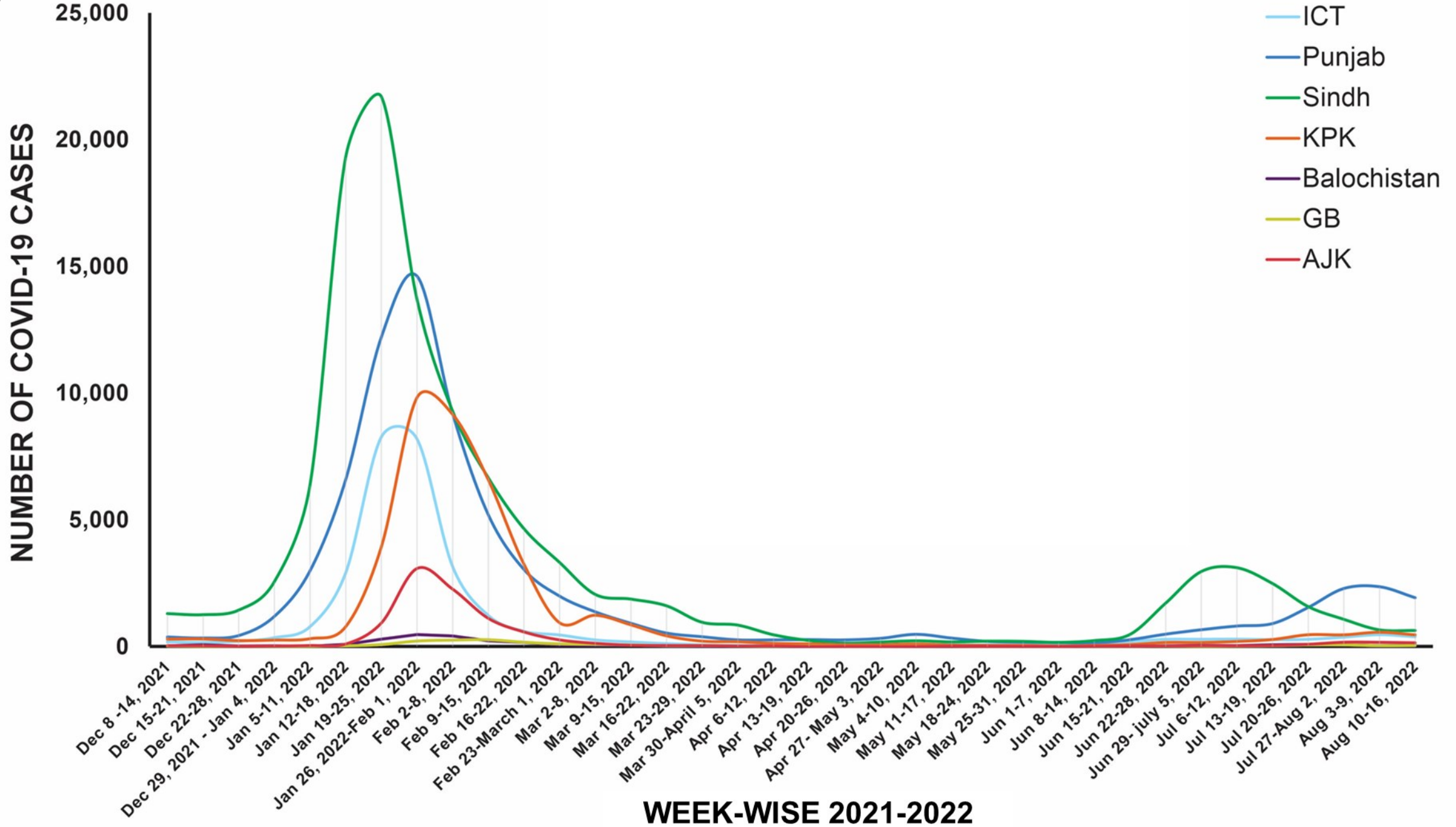


Figure 3

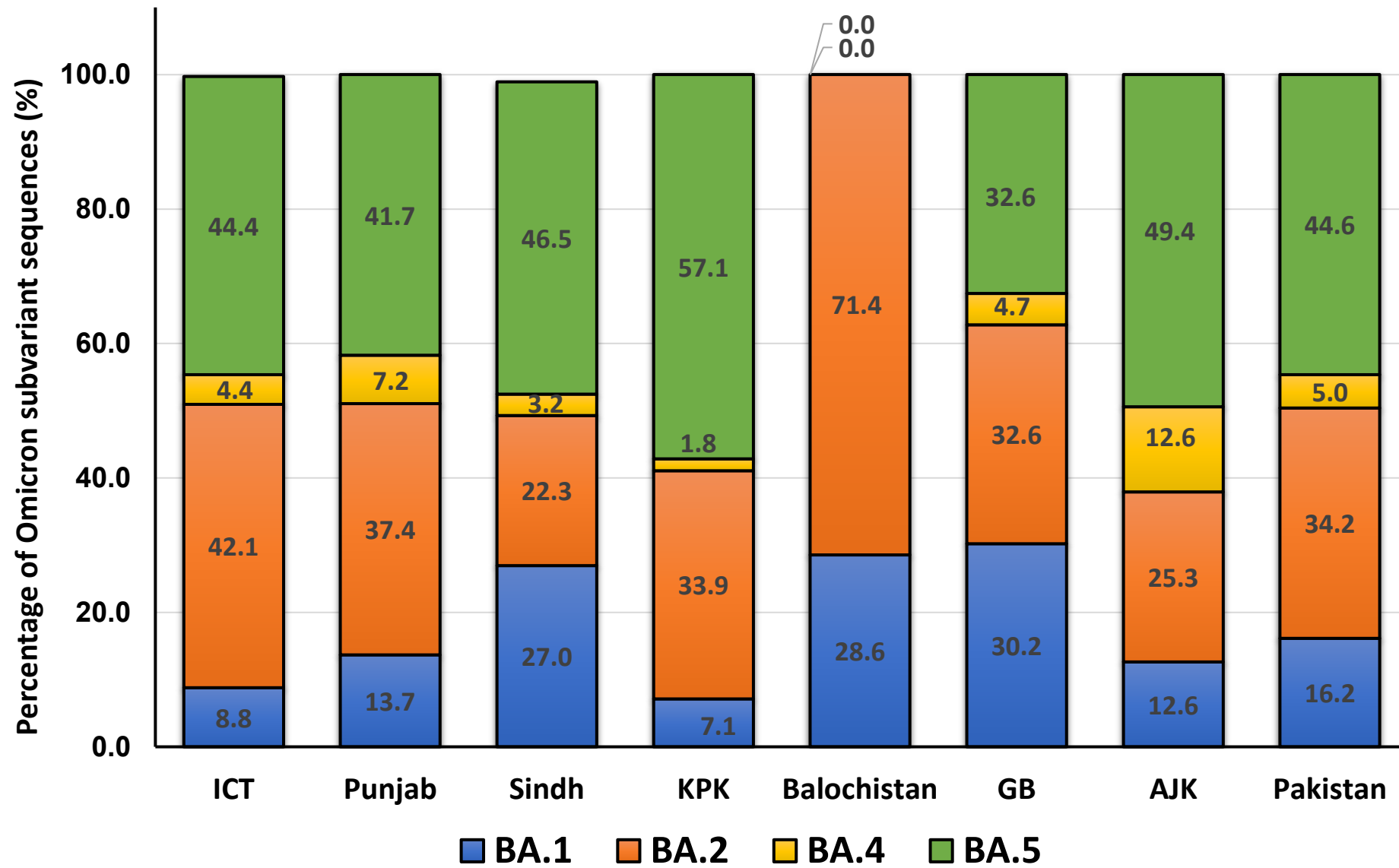


Figure 4

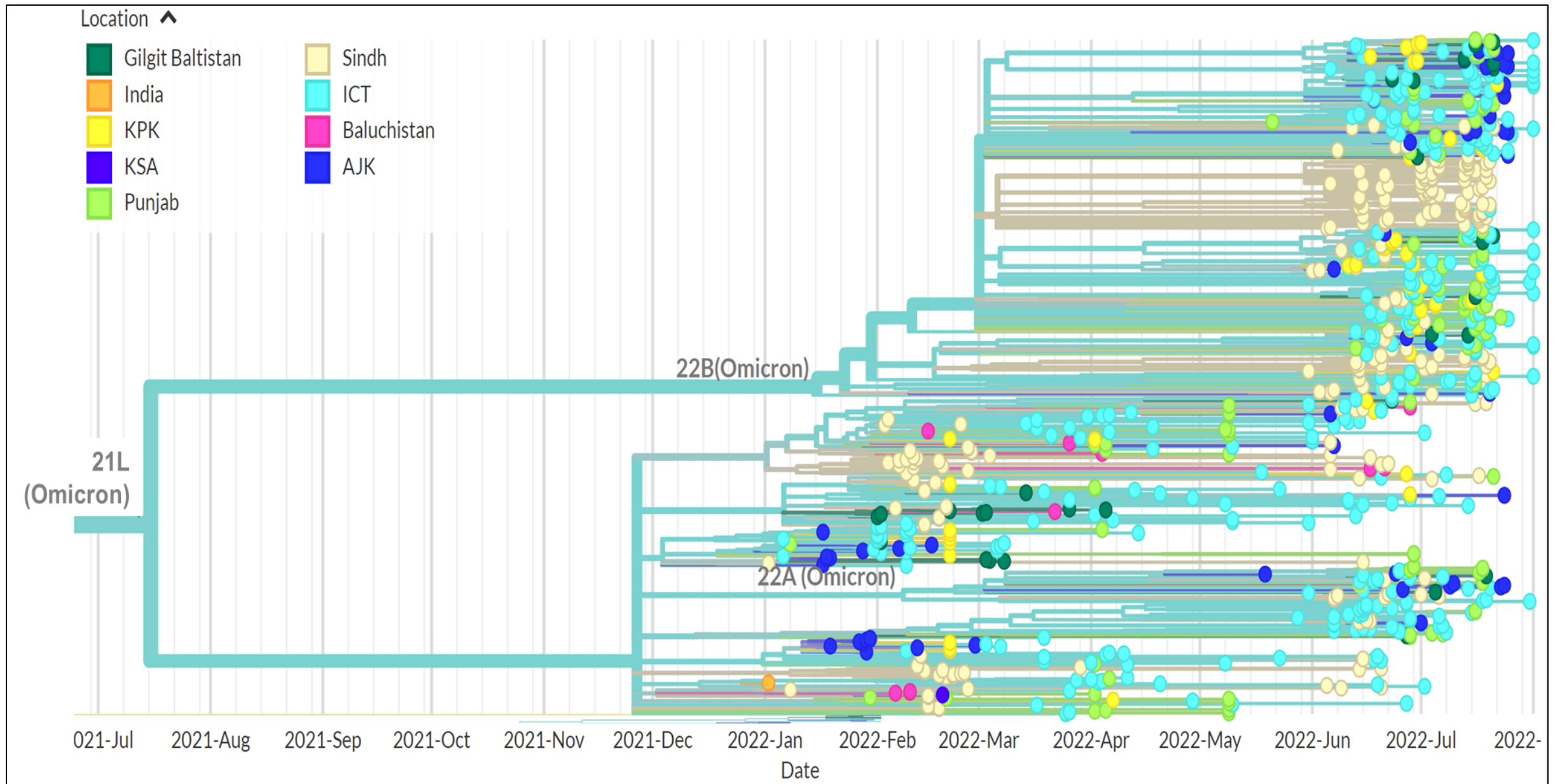


Figure 5A

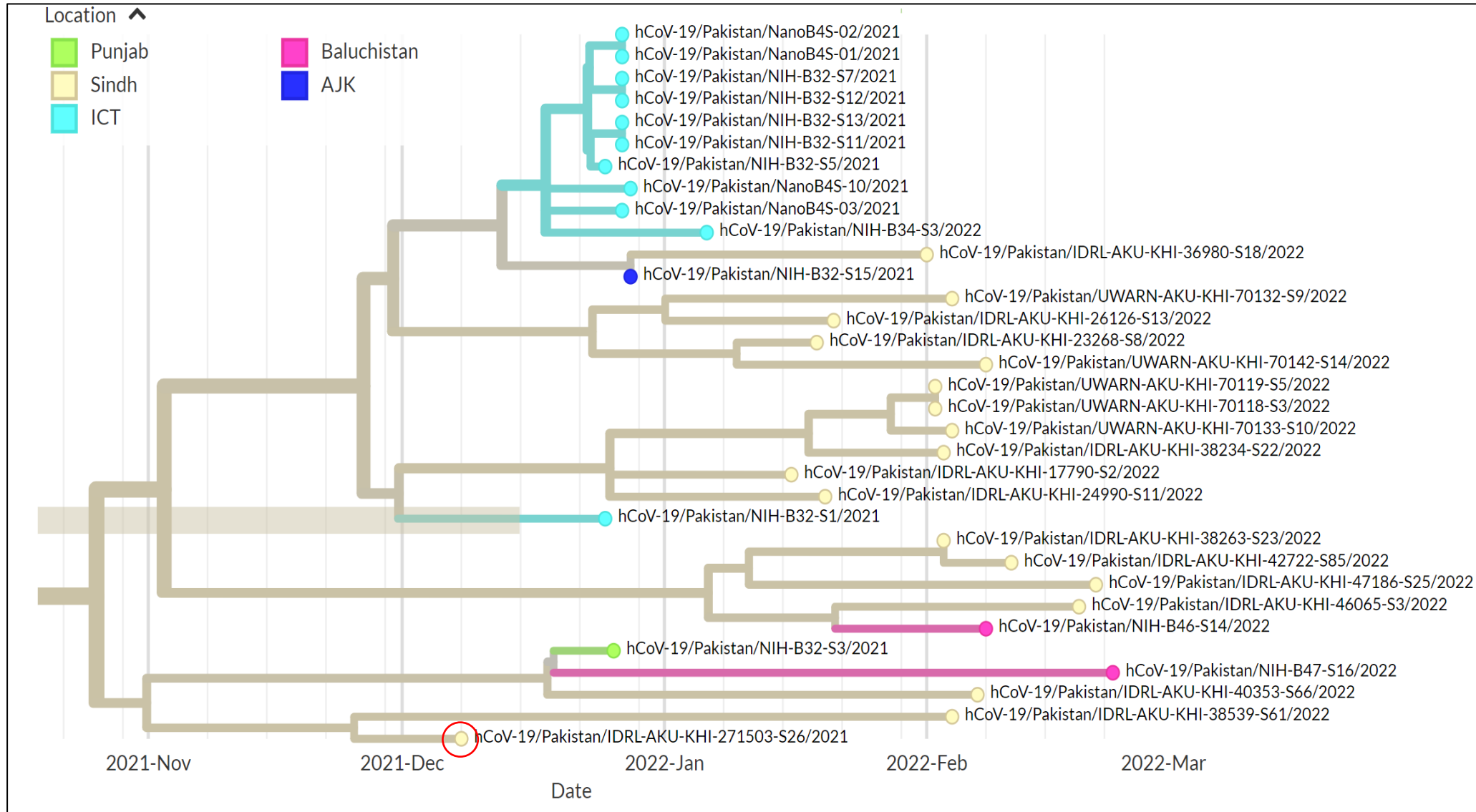


Figure 5B

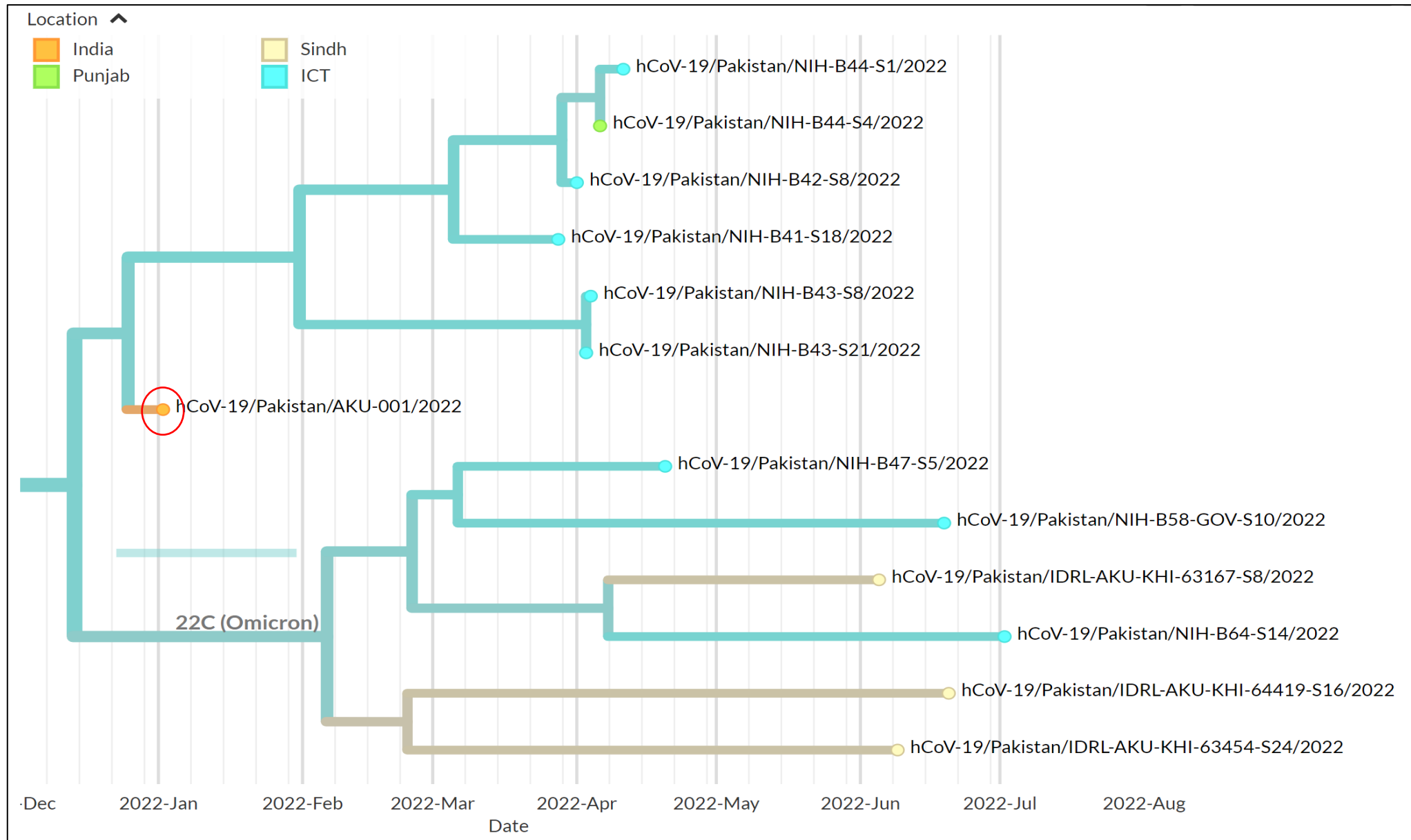


Figure 5C

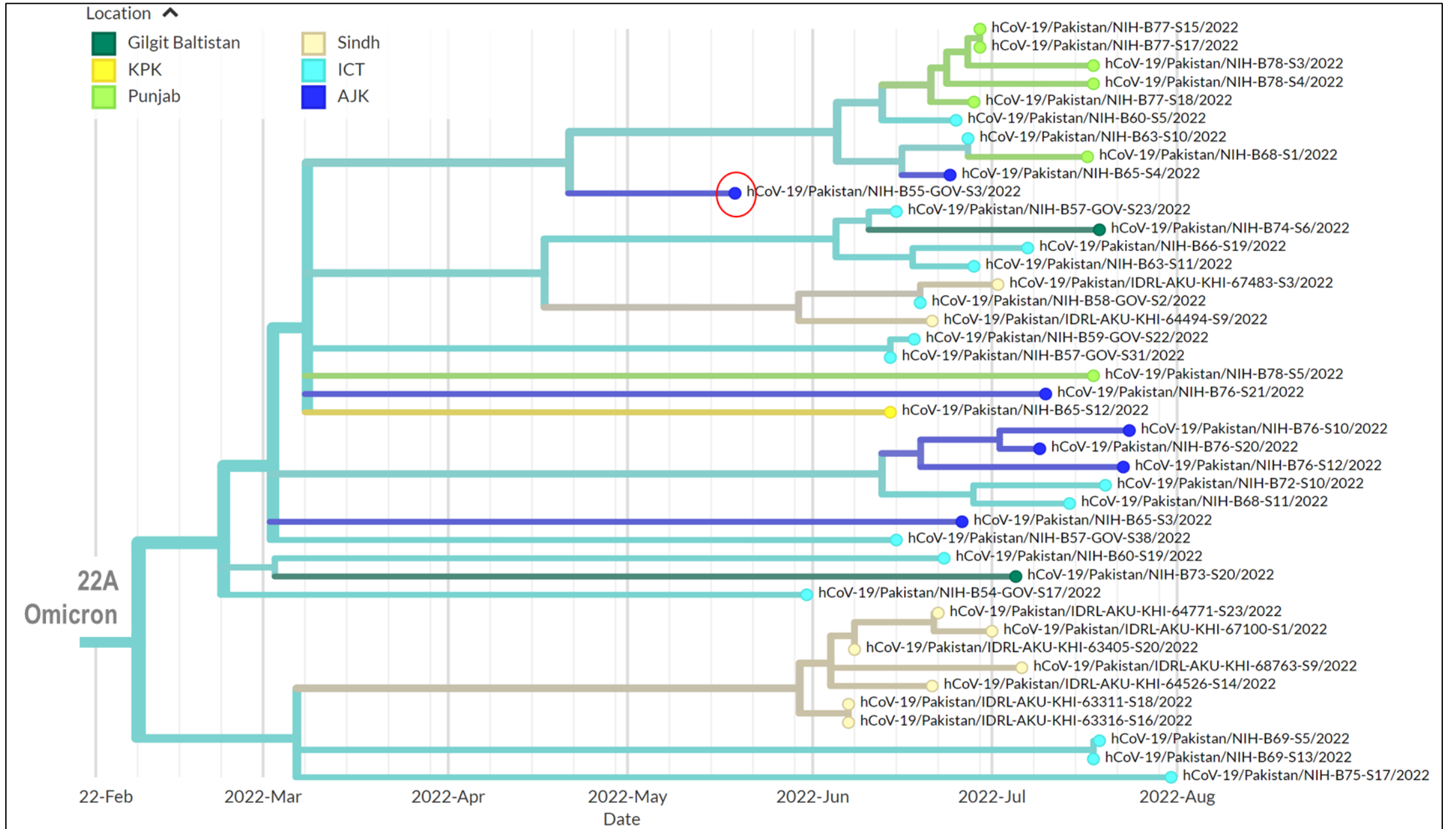


Figure 5D

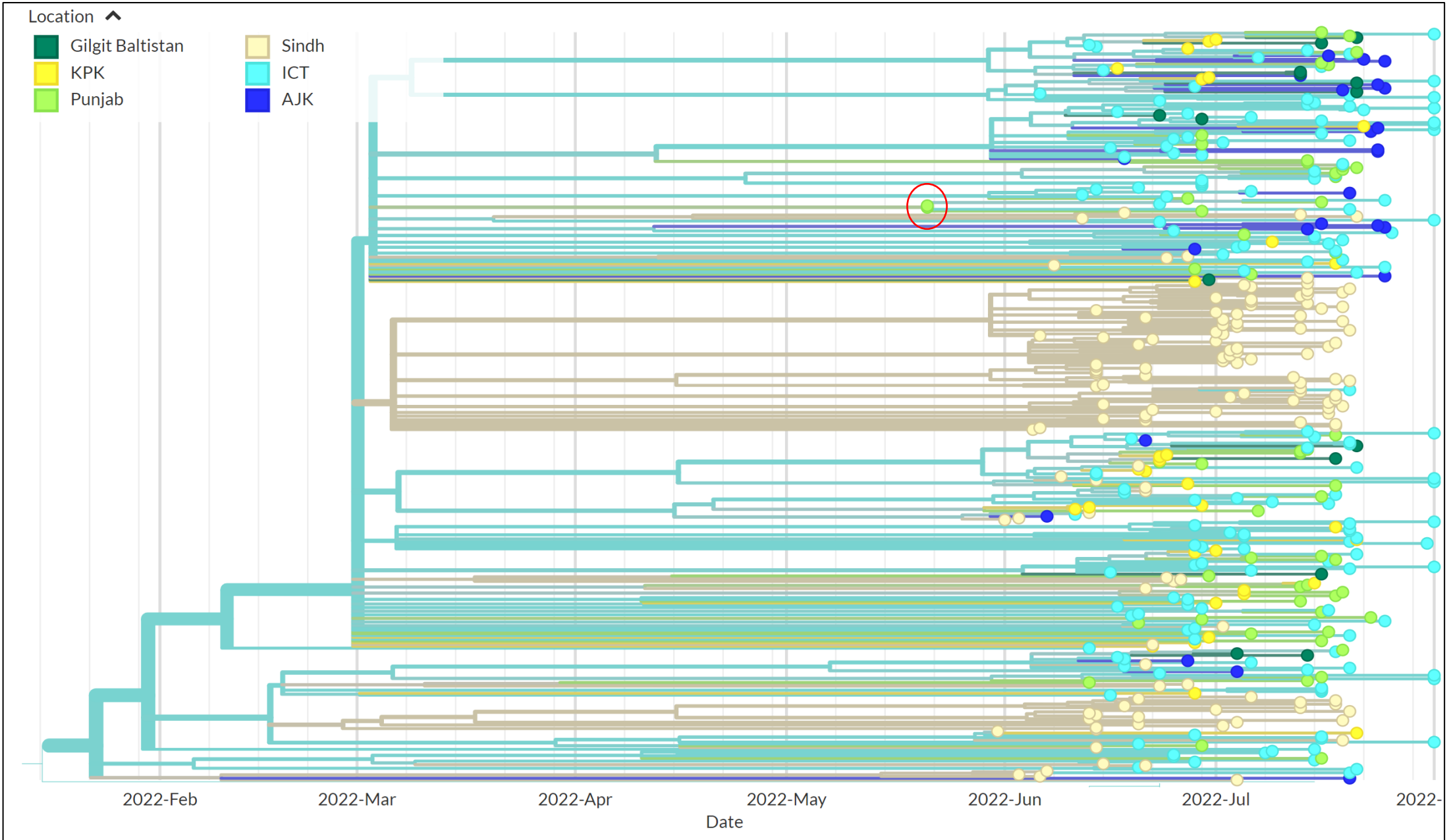
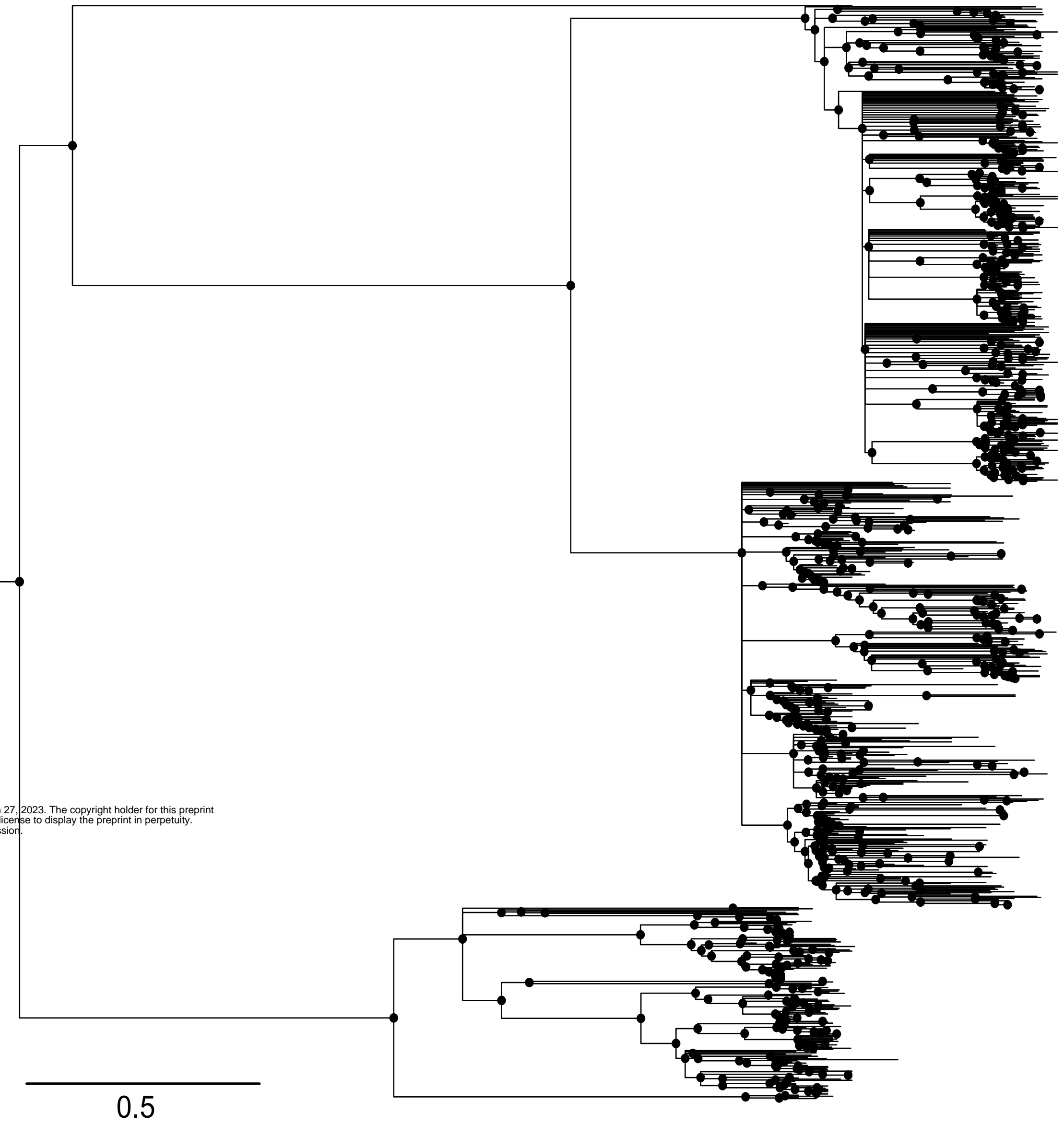


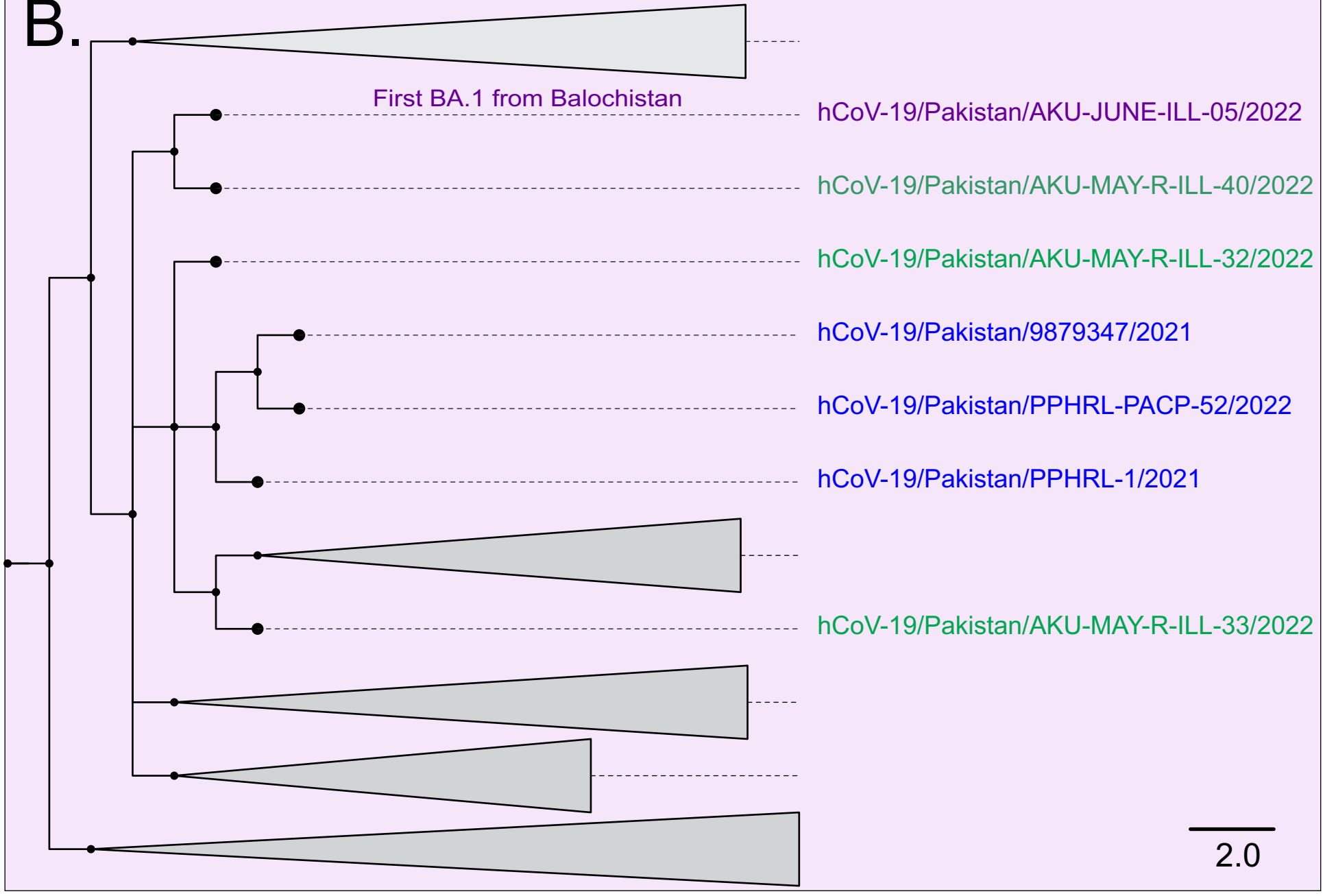
Figure 6

A.

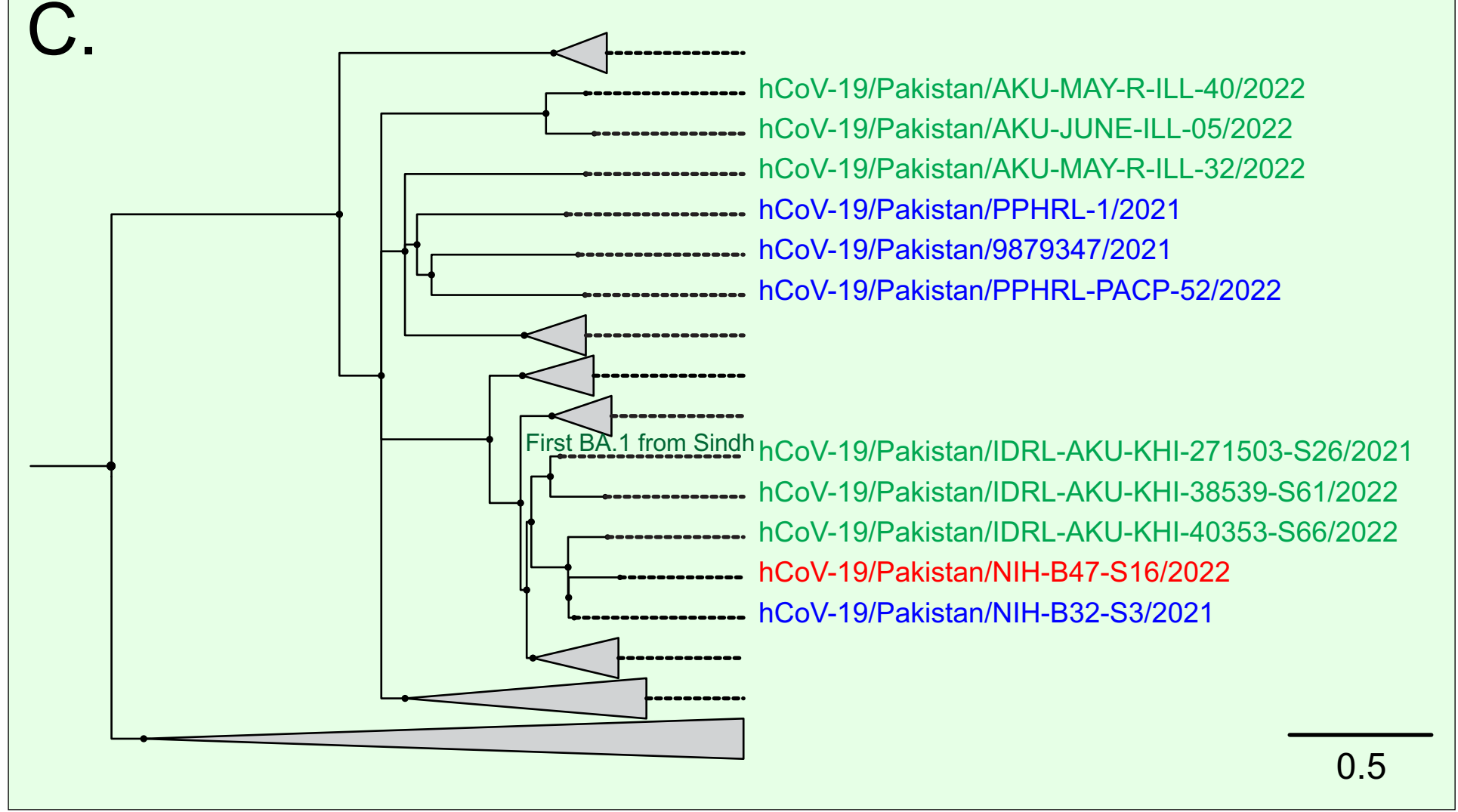


medRxiv preprint doi: <https://doi.org/10.1101/2023.03.25.23287718>; this version posted March 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

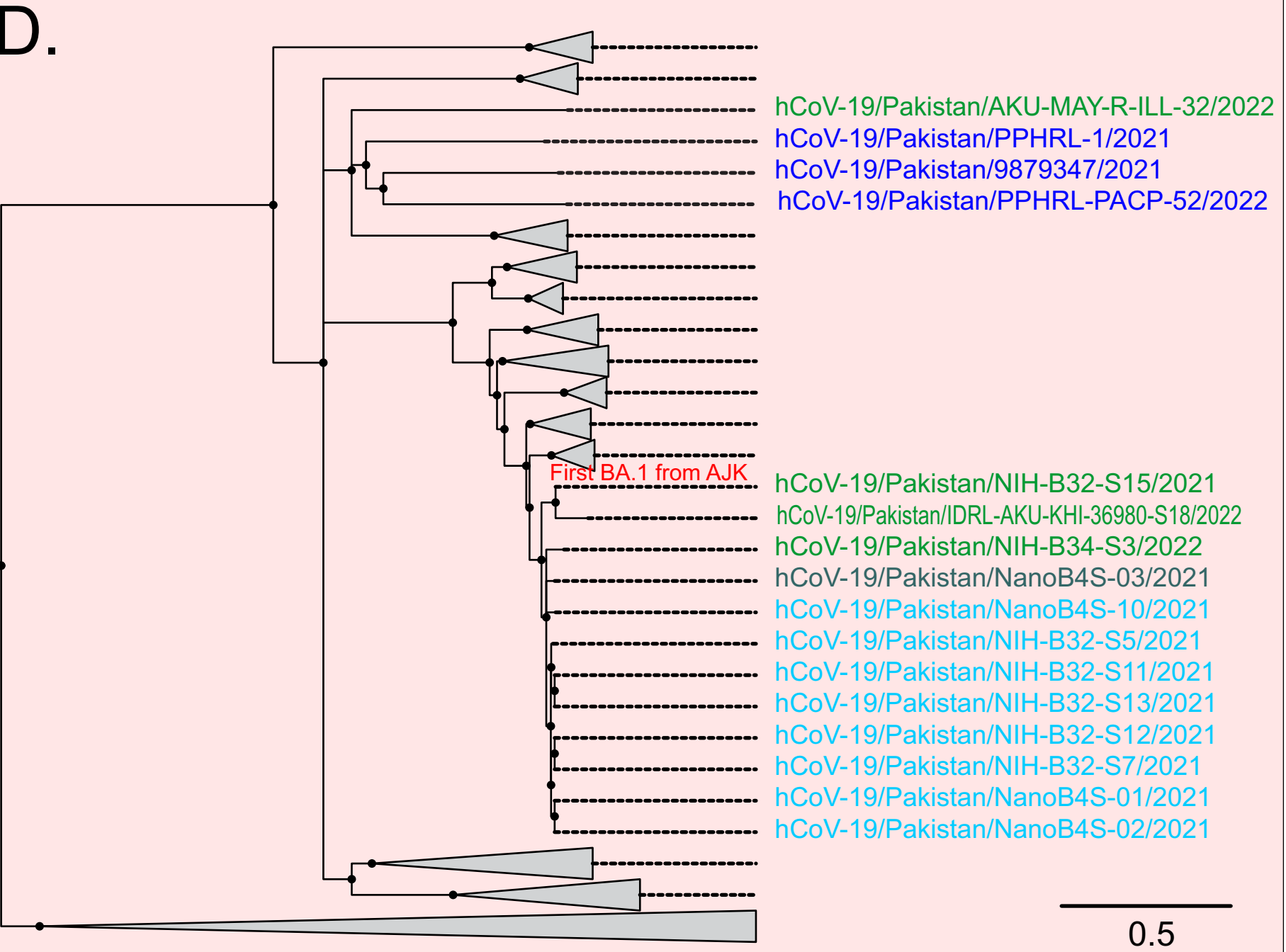
B.



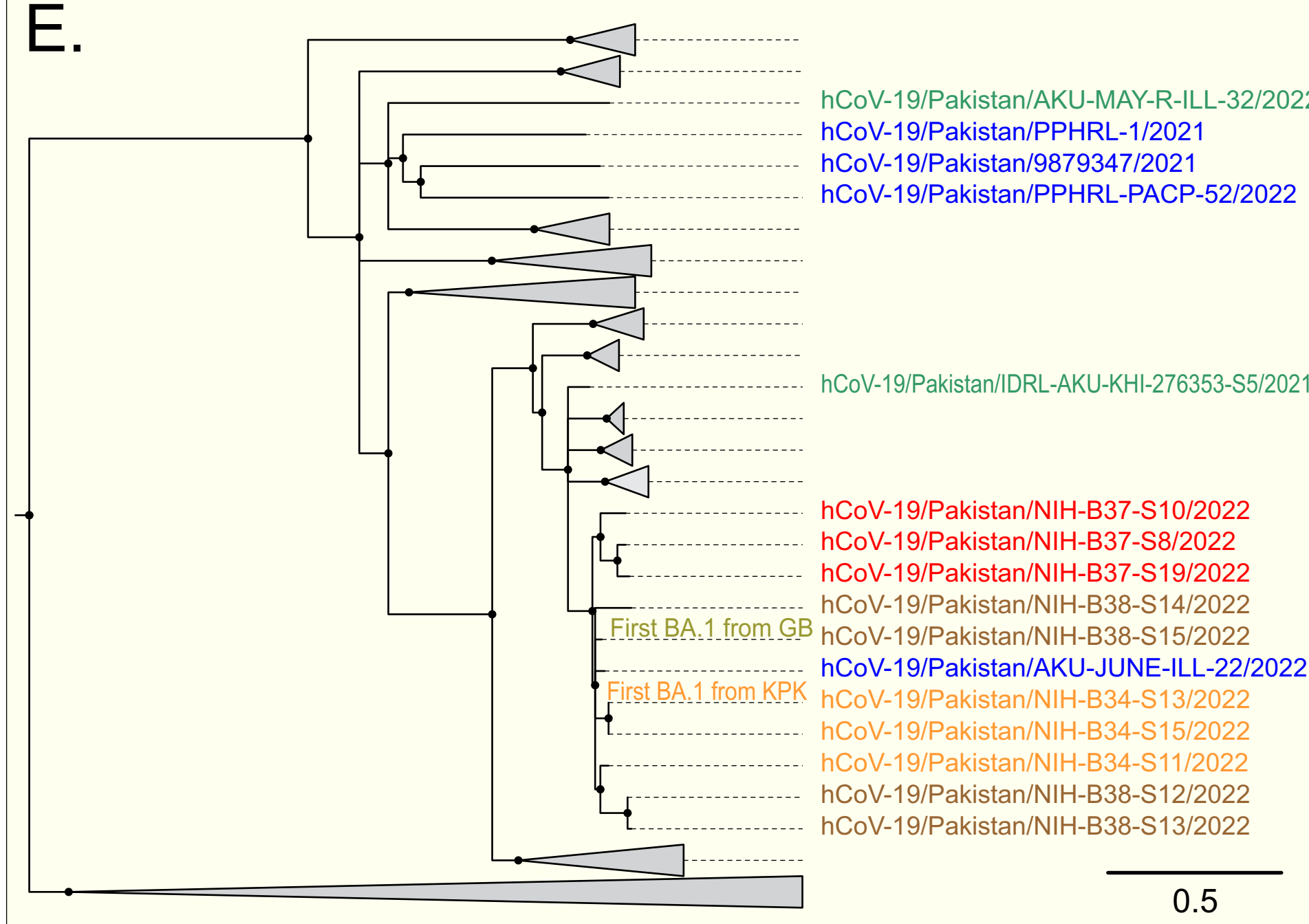
C.



D.



E.



F.

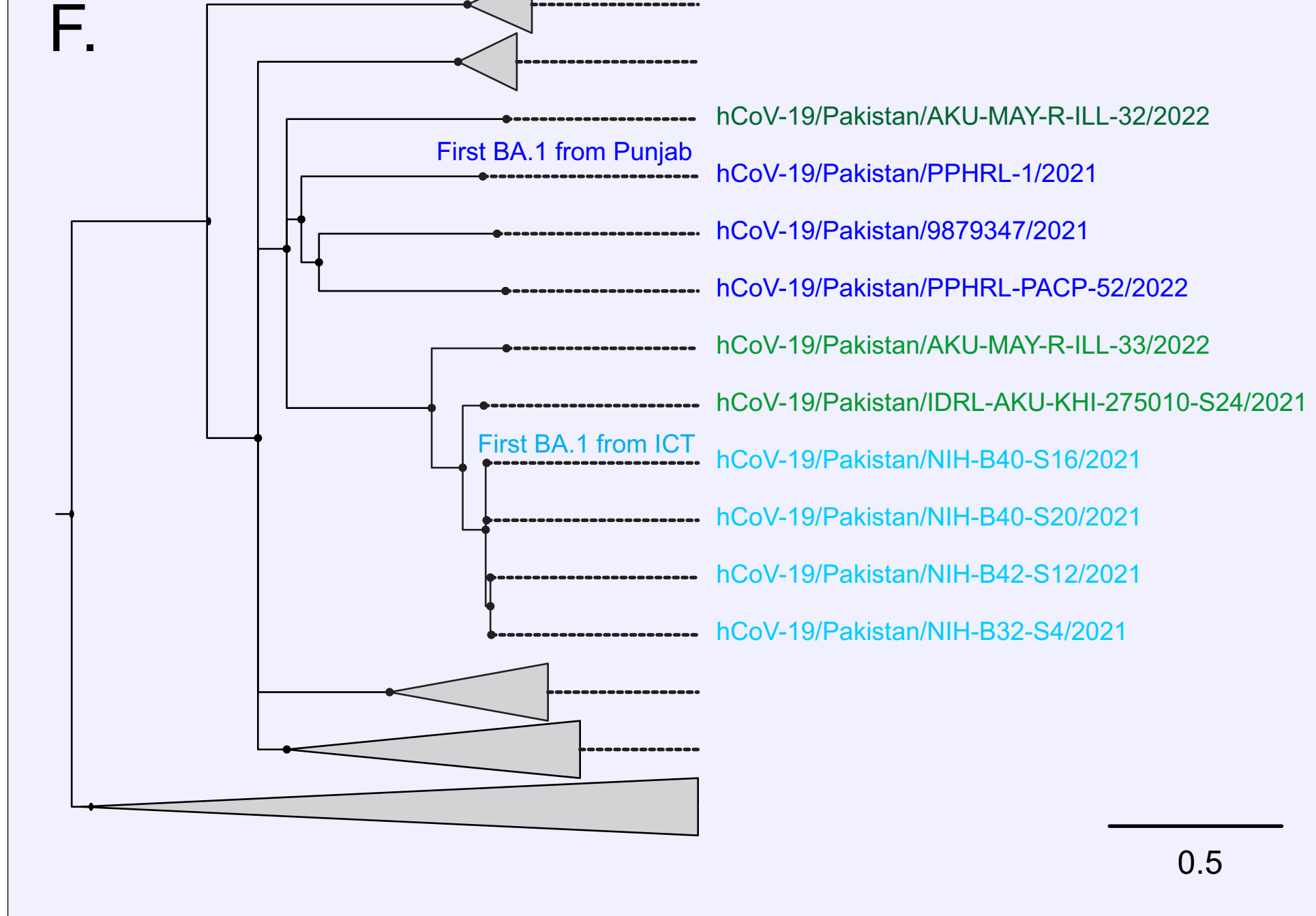


Table 1. Distribution of SARS-CoV2 lineages according to different age groups. The age-group wise SARS-CoV2 variants frequency is presented. The Kruskal-Wallis test was performed to study association between the age groups for every variant and p-value <0.05 is considered significant.

SARS-CoV2 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, 1.7%)	≤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=161, 16.6%)	≤18	12	7.5	<0.0001
	19-40	71	44.1	
	41-55	31	19.3	
	≥ 56	47	29.2	
BA.2 (n=312, 32.2%)	≤18	34	10.9	<0.0001
	19-40	153	49.0	
	41-55	58	18.6	
	≥ 56	67	21.5	
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=391, 40.3%)	≤18	40	10.2	<0.0001
	19-40	199	50.9	
	41-55	68	17.4	
	≥ 56	84	21.5	