

1 **Genomic epidemiology of SARS-CoV-2 divulge B.1,**
2 **B.1.36, and B.1.1.7 as the most dominant lineages in**
3 **first, second, and third wave of SARS-CoV-2**
4 **infections in Pakistan**

5

6

7 Atia Basheer and Imran Zahoor*

8

9 Genetics and Genomic Laboratory, Department of Animal Breeding and Genetics, University
10 of Veterinary and Animal Sciences, Ravi Campus, Pattoki, Punjab, Pakistan

11

12

13

14

15

16

17

18

19

20

21

22

23 *Corresponding author

24 imran.zahoor@uvas.edu.pk

25

26 **Abstract**

27 The present study aims to investigate the genomic variability and epidemiology of SARS-CoV-2
28 in Pakistan along with their role in the spread and severity of infection during the three waves of
29 COVID-19. A total of 453 genomic sequences of Pakistani SARS-CoV-2 were retrieved from
30 GISAID and subjected to MAFFT-based alignment and QC check which resulted in removal of
31 53 samples. The remaining 400 samples were subjected to Pangolin-based genomic lineage
32 identification. And to infer our SARS-CoV-2 time-scaled and divergence phylogenetic trees,
33 3,804 selected global reference sequences plus 400 Pakistani samples were used for the
34 Nextstrain analysis with Wuhan/Hu-1/2019, as reference genome. Finally, maximum likelihood
35 based phylogenetic tree was built by using the Nextstrain & coverage map was created by
36 employing Nextclade. And by using the amino acid substitutions the maximum likelihood
37 phylogenetic trees were developed for each wave, separately. Our results reveal the circulation of
38 29 lineages, belonging to following 7 clades G, GH, GR, GRY, L, O, & S in the three waves.
39 From first wave, 16 genomic lineages of SARS-CoV-2 were identified with B.1(24.7%),
40 B.1.36(18.8%), & B.1.471(18.8%) as the most prevalent lineages respectively. The second wave
41 data showed 18 lineages, 10 of which were overlapping with the first wave suggesting that those
42 variants could not be contained during the first wave. In this wave, a new lineage, AE.4, was
43 reported from Pakistan for the very first time in the world. However, B.1.36 (17.8%), B.1.36.31
44 (11.9%), B.1.1.7 (8.5%) & B.1.1.1 (5.9%) were the major lineages in second wave. Third wave
45 data showed the presence of 9 lineages with Alpha/B.1.1.7 (72.7%), Beta/B.1.351 (12.99%), &
46 Delta/B.1.617.2 (10.39%) as the most predominant variants. It is suggested that these VOCs
47 should be contained at the earliest in order to prevent any devastating outbreak of SARS-CoV-2
48 in the country.

49 **Keywords:** SARS-CoV-2; genomic epidemiology; lineages; clades; phylogenetic analysis;

50 Pakistan

51

52 Introduction

53 Since the initial reports of coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China on
54 30 December 2019 [1, 2], SARS-CoV-2 has been spreading worldwide and, as of 15 July 2021,
55 there have been 188.13 million confirmed infections and 4.06 million deaths [3]. This was the
56 reason that SARS-CoV-2 epidemic was declared a global pandemic by World Health
57 Organization (WHO) on 11 March 2020 [4]. SARS-CoV-2 is a single-stranded positive-sense
58 RNA virus belonging to the genus *Betacoronavirus*, and subgenus *Sarbecovirus*. The genome
59 size of SARS-CoV-2 is approximately 30kb and its genomic structure has followed the
60 characteristics of known genes of the coronavirus [5]. Albeit, RNA viruses are highly prone to
61 mutations but Coronaviruses (CoVs) have an outstanding and distinctive feature of intrinsic
62 proofreading mechanism [6]. However, CoVs encode a protein called nonstructural protein 14
63 (nsp14) that possesses a 3' to 5' exonuclease (ExoN) activity which is necessary for replication
64 fidelity and proofreading activity [7]. Owing to the presence of large and complex genome, this
65 proofreading mechanism is considered critical for maintaining normal functioning and fitness of
66 CoVs. And, perhaps, this is the reason that mutation rate in SARS-CoV-2 is 10-fold lesser
67 compared with other RNA viruses. However, even then geneticists had reported a rate of 33
68 mutations/year in SARS-CoV-2 genome and now scientists are using these mutations to
69 categorize different variants of this virus into different clades, lineages, and sublineages [4]. To
70 date the SARS-CoV-2 has been divided in more than 81 lineages, on the basis of variations in its
71 genome, which vary significantly in their transmissibility and virulence [8]. All the existing
72 lineages including variants of concerns are the descendants of two ancestral lineages, A and B,
73 discovered from China in the start of the pandemic. However, European Center for Disease
74 Prevention and Control (ECDC) has demonstrated five variants of concerns (VOCs) due to their

75 high transmissibility, pathogenicity, and effects on the vaccine efficacy and that's why they are
76 known as VOCs. Alpha variant (B.1.1.7) also known as Alpha VOC, was first detected in UK in
77 September 2020 [9]. This variant is highly variable and contain more than a dozen mutations
78 compared with wild-type lineages. It had also been reported to have 50-70% high transmissibility
79 [10], greater severity of the disease [11, 12], and effects on the efficacy of the vaccine compared
80 with wild-type lineage. Beta variant of concern, B.1.351, also known as 20H/501Y.V2 was
81 identified from South Africa. It shares several mutations with B.1.1.7 and has the following
82 major amino acid mutations, K417N, E484K, N501Y, D614G, & A701V in its spike
83 protein. This lineage is also reported to have increased transmissibility and even have the
84 advantage of escaping from immunity due to the presence of E484K mutation in its spike protein
85 [13]. Likewise, B.1.617.2 lineage (Delta variant), also known as 20A/S:478K, was identified in
86 India at the end of 2020 and had been reported to have the following major spike mutations,
87 L452R, T478K, D614G, & P681R, which are involved in enhancing its transmissibility and risk
88 of hospitalization compared with B.1.1.7 [14-16]. Lineage B.1.617.2, also known as Delta
89 variant of concern was detected from India in December 2020 with following mutation of
90 interest L452R, T478K, D614G, P681R in its spike protein. Likewise, P.1 lineage (Gamma), also
91 known as 20J/501Y.V3, was identified from Brazil [17]. It has K417T, E484K, N501Y,
92 D614G, and H655Y as the most noteworthy amino acid mutations which have impact on its
93 transmissibility and immunity [18]. B.1.427 and B.1.429 lineages (Epsilon), also collectively
94 known as 20C/S452R [19] are two more variants of concern which are based on several spike
95 protein mutations, including L452R, which is associated with increased cell entry and reduced
96 susceptibility to neutralization by convalescent and vaccine recipient plasma in vitro [20].

97 In Pakistan, first two cases of COVID-19 were confirmed on February 26, 2020 and by June 17,
98 2020 each region of Pakistan has reported at least one confirmed case of Coronavirus disease.
99 Pakistan has experienced three different waves of COVID-19, to date. The first nationwide
100 COVID-19 wave started in late May 2020 and reached its peak in mid-June, when the number of
101 newly confirmed cases and the number of daily deaths reached at its peak, and then in mid-July,
102 it ended abruptly. The second wave of COVID-19 started in early November 2020 but the
103 intensity of this wave was relatively low, and it mainly affected the southern part of Sindh. The
104 country's third wave began in mid-March 2021 and it mainly affected the Punjab and Khyber
105 Pakhtunkhwa provinces. This wave peaked in late April 2021, and since then, number of new
106 cases and daily deaths have been decreasing. In Pakistan, 978,662 individuals have been infected
107 with SARS-CoV-2 and 22,642 of them have died until 15 June 2021. The first three cases of
108 B.1.1.7 (UK-VOC) were reported on 26 February 2021, and perhaps this was the reason that the
109 third wave of COVID-19 was much severe compared with first two waves. Hence, it is
110 hypothesized that multiple variants of SARS-CoV-2 are present in Pakistan which are perhaps
111 the reason for the repeated occurrence of COVID-19 epidemics in the country. However, the
112 scarcity of genomics and epidemiologic studies on the transmission, spread, and distribution of
113 different variants of virus in Pakistan has further hampered the control efforts in the country.
114 Thus, the aims of the current study are to identify the lineages along with their origin,
115 transmission, and prevalence in the country and the estimation of genetic relatedness among the
116 variants, present in Pakistan and world, through phylogenetic clustering by using the whole
117 genome sequence data.

118

119 **Results**

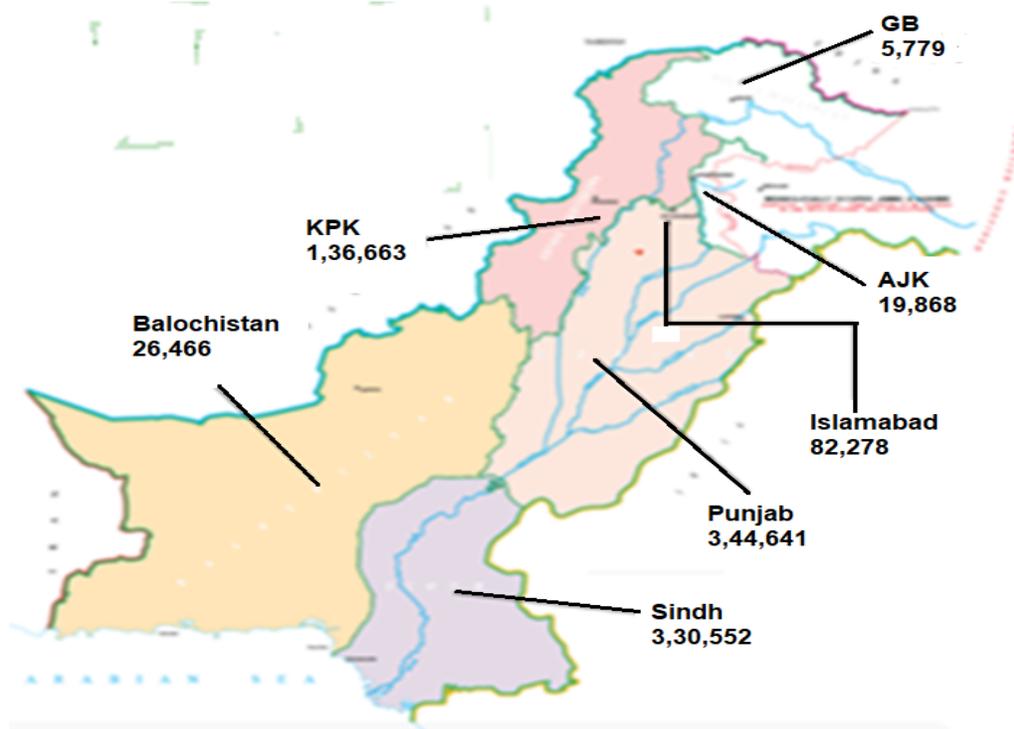
120 The data about the health status of COVID-19 patients in different geographical regions of
121 Pakistan, taken from official website of Pakistan (www.covid.gov.pk), is presented in Table 1
122 and Fig 1.

Areas	Diseased	Hospitalized	Deceased	Recovered
Azad Jammu & Kashmir (AJK)	19,868	447	571	18850
Baluchistan	26,446	794	300	25372
Gilgit-Baltistan	5,759	105	108	5546
Islamabad	82,278	1216	773	80289
Khyber Pakhtoon Khawah (KPK)	136,663	2958	4252	129453
Punjab	344,641	10462	10603	323576
Sindh	330,552	19827	5306	305419

123
124 **Table 1** The data of COVID-19 taken from official covid website (<http://www.covid.gov.pk>) of
125 Pakistan for a period spanning from 3 January 2020 to 16 June 2021.

126
127
128 It is observed that maximum COVID-19 cases were detected in Punjab (344,641) followed by
129 Sindh (330,552), KPK (136,663), Baluchistan (26,446), AJK (19,868), Capital city of
130 Islamabad (82,278) and Gilgit-Baltistan (5,759) region of Pakistan (Fig. 1). During this
131 duration, 21,913 deaths were recorded in Pakistan, out of which the highest numbers (10,603)
132 were observed in Punjab province.

133
134
135



136

137

138 **Fig. 1** Number of the total cases of COVID-19 in different provinces/territories of Pakistan up to
139 16 June 2021. COVID cases of Islamabad, Capital of Pakistan, are also shown here. This map is
140 adopted from <https://covid.gov.pk/>.

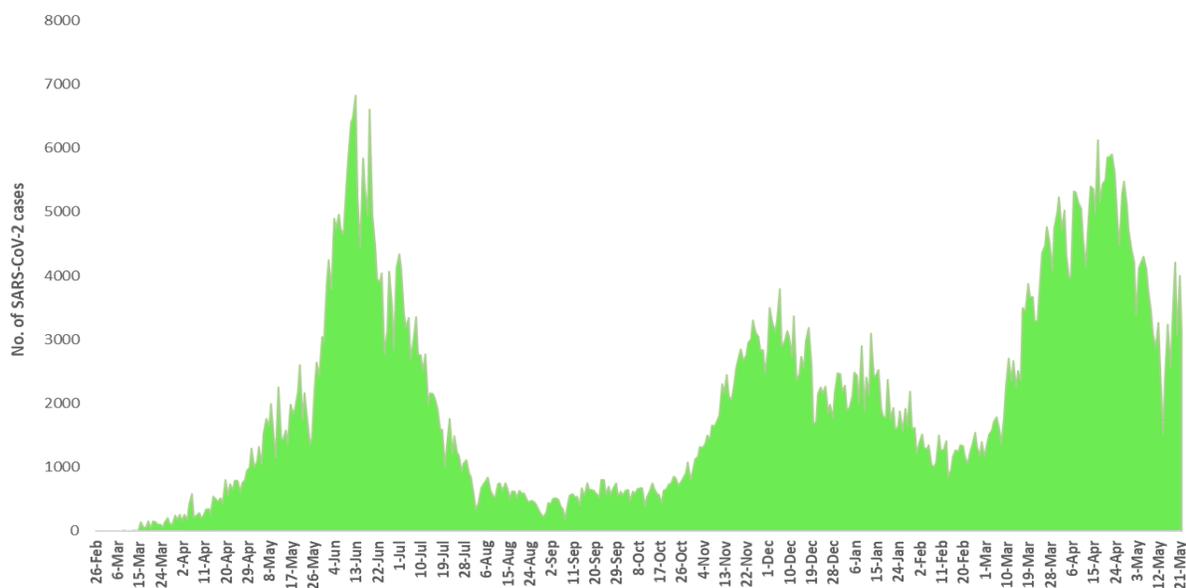
141

142

143 **Major waves of SARS-CoV-2 infection in Pakistan**

144 Pakistan has observed three major waves of SARS-CoV-2 infection. The nation's first wave of
145 SARS-CoV-2 began in late May 2020, peaked in mid-June and ended in mid-July. In this wave,
146 Pakistan experienced the highest number of cases and mortalities during its peak month.
147 However, this wave passed away very rapidly and new cases and mortalities were dropped
148 suddenly after the peak. Second surge of COVID-19 started from end of October 2020 to mid-

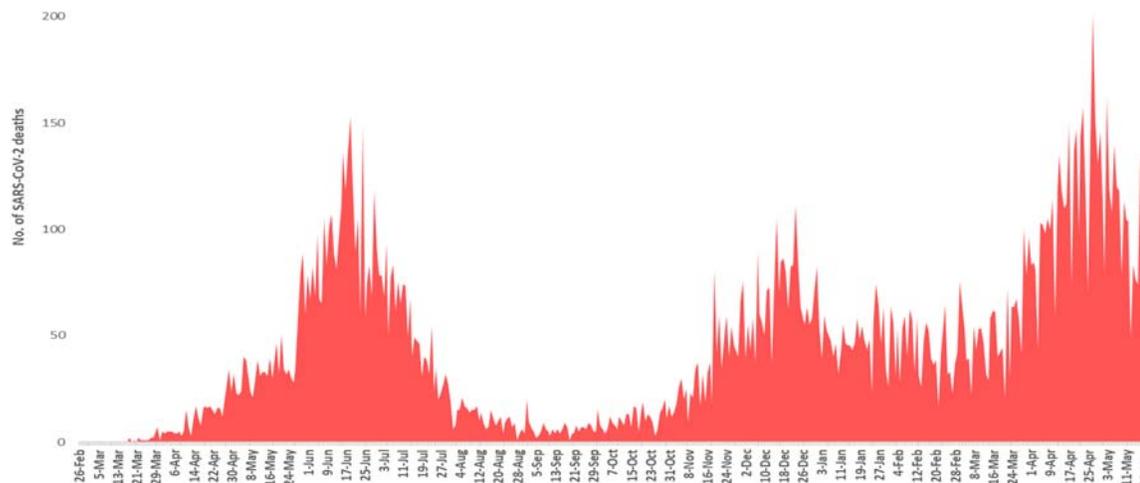
149 February 2021. This wave was low in its intensity, mainly affecting the southern parts of Sindh
150 province, and peaked in mid-December 2020. The country's third wave began in mid-March
151 2021, when daily new confirmed cases and mortalities began to rise steeply. The third wave
152 mainly affected the provinces of Punjab and Khyber Pakhtunkhwa. This wave peaked in late
153 April 2021, and since then daily new cases, and new mortalities have been falling. The
154 complete pictures of number of SARS-CoV-2 positive cases and mortalities from February
155 2020 to May 2021 is shown in Fig. 2 and 3.



156

157

Fig 2. Month-wise details of SARS-CoV-2 cases starting from February 2020 to May 2021

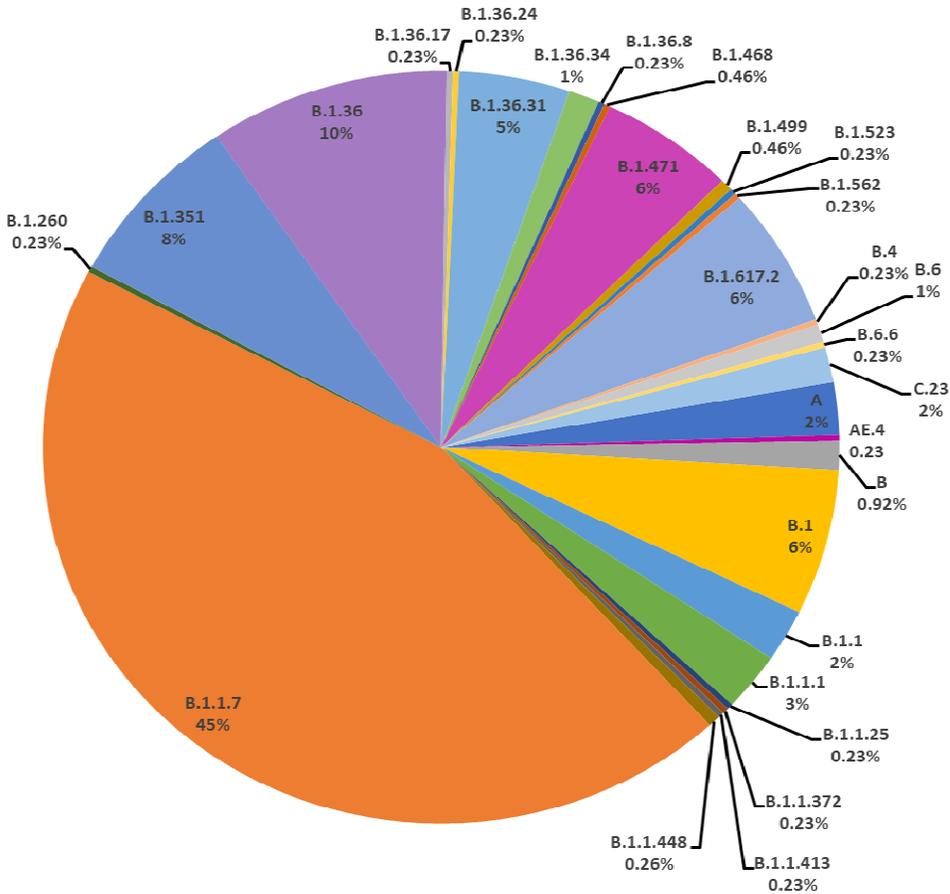


158

159 Fig 3. Month-wise details of deaths caused by SARS-CoV-2 starting from February 2020 to May
160 2021

161 **Distribution of Genomic Lineage of SARS-CoV-2 in Pakistan**

162 In the current study, a total of 453 SARS-CoV-2 genomes sequences downloaded from GISAID
163 on 10 July 2021 were used. In Pakistan, 29 lineages were observed by using Phylogenetic
164 Assignment of Named Global Outbreak Lineage (<http://www.pangolin.cog-uk.io>). Out of total,
165 53 samples were failed. The overall lineage distribution highlighted the occurrence of
166 B.1.1.7(n=178), B.1.36(n=39), B.1.351(n=30), B.1(n=25), B.1.617.2(n=24), B.1.471(n=22),
167 B.1.36.31(n=18), B.1.1.1(n=10)), B.1.1(n=9), A(n=9), C.23(n=6), B(n=5), B.1.36.17(n=1),
168 B.1.36.34(n=5), B.6(n=3), B.1.499(n=2), B.1.1.448(n=2), AE.4(n=1), B.1.260(n=1),
169 B.1.36.24(n=1), B.1.36.8(n=1), B.1.468(n=1), B.1.523(n=1), B.1.562(n=1), B.1.1.25(n=1),
170 B.1.1.372(n=1), B.1.1.413(n=1), B.4(n=1) and B.6.6(n=1) as mentioned in Fig. 4 and Table 2.



171

172 Fig. 4. All the lineages observed from Pakistan sequences downloaded on 10 July 2021.

173 **Table 2. List of 29 lineages of SARS-CoV-2 observed in Pakistan along with their local and**
 174 **global percentages**

175

Clade	Pakistan Lineages	No of samples sequences from Pakistan	Percentage of Lineages in Pakistan	No of samples sequenced globally of Pakistani Lineages	Global Percentage of Pakistani Lineages
G	B.1.617.2	24	6	158,368	0.1178
GH	B.1	25	6.25	85,731	0.0638
	B.1.351	30	7.5	28,385	0.0211
	B.1.36	39	9.75	6,059	0.0045
	B.1.36.17	1	0.25	2027	0.0015
	B.1.36.24	1	0.25	158	0.0001
	B.1.36.31	18	4.5	543	0.0004
	B.1.36.34	5	1.25	43	0.0000
	B.1.36.8	1	0.25	936	0.0007
	B.1.468	1	0.25	172	0.0001
	B.1.471	22	5.5	350	0.0003
B.1.499	2	0.5	760	0.0006	

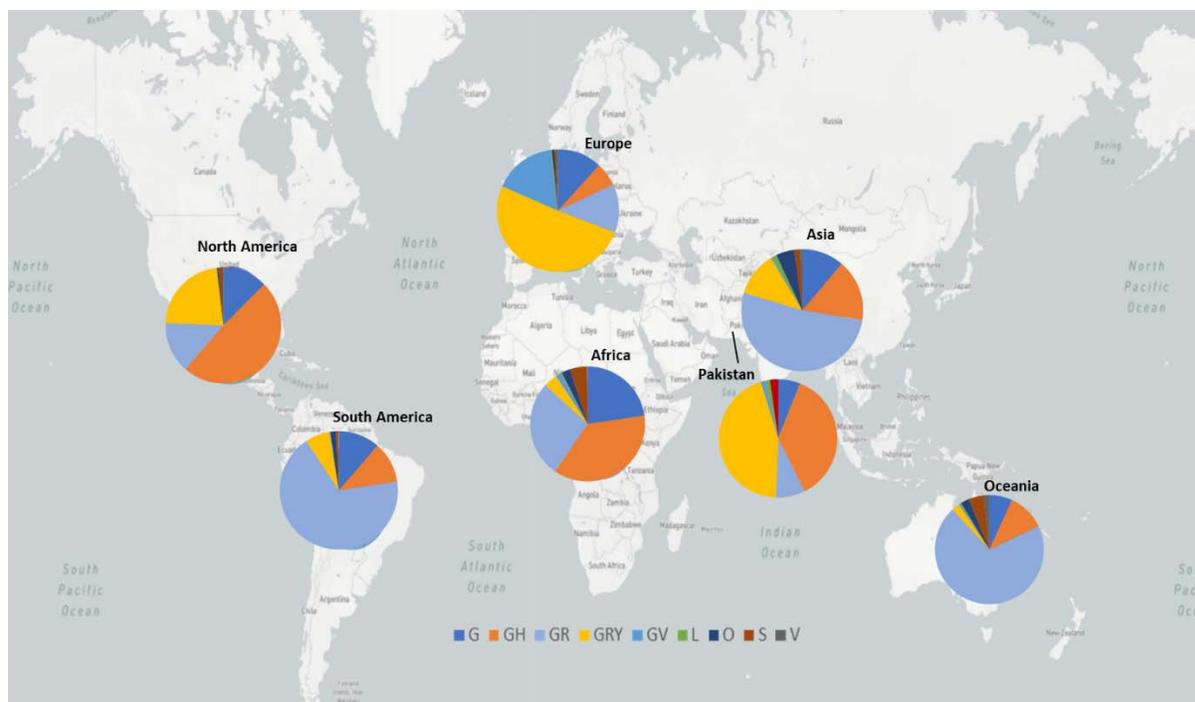
	B.1.523	1	0.25	388	0.0003
	B.1.562	1	0.25	53	0.0000
GR	AE.4	1	0.25	20	0.0000
	B.1.1	9	2.25	48,427	0.0360
	B.1.1.1	10	2.5	2774	0.0021
	B.1.1.25	1	0.25	1641	0.0012
	B.1.1.372	1	0.25	1339	0.0010
	B.1.1.413	1	0.25	105	0.0001
	B.1.1.448	2	0.5	468	0.0003
	C.23	6	1.5	153	0.0001
GRY	B.1.1.7	178	44.5	994,397	0.7395
L	B	5	1.25	5918	0.0044
	B.4	1	0.25	439	0.0003
O	B.1.260	1	0.25	168	0.0001
	B.6	3	0.75	1052	0.0008
	B.6.6	1	0.25	1043	0.0008
S	A	9	2.25	2694	0.0020
Total Pakistani sequences analyzed		400	No. of total global sequences	2,302,116	

176

177

178 **Distribution of GISAID clades of SARS-CoV-2 in Paksitan**

179 The distribution of GISAID clades of SARS-CoV-2 genomes from Pakistani & global dataset,
 180 downloaded on 10 July 2021, is presented in Table 2 & Fig 5. The genomic sequences of
 181 Pakistani SARS-CoV-2 samples indicate the prevalence of 7 clades with following order of
 182 prevalence GRY (44.5%), GH (36.75%), GR (7.75%), G (6%), S (2.25%), L (1.50%), & O
 183 (1.25%).



184

185 **Fig 5. GISAID clades of SARS-COV-2 genome from global and Paksitani datasets. The map in the**
186 **background is adopted from <https://nextstrain.org/sars-cov-2/> [21]**

187

188 **Occurrence of Variants of concerns in Pakistan**

189 Our data also revealed the presence of the three VOCs including Alpha (B.1.17), Beta (1. 351)
190 and Delta (B.1.617.2) variants in Pakistan, with the following order of genomic prevalence
191 44.5%, 7.5% and 6% respectively.

192

193 **Phylogenetic Analysis**

194 Maximum Likelihood Phylogenetic tree of 400 whole genome sequences of Pakistani SARS-
195 CoV-2 was developed as per definitions of the PANGOLIN lineage and GISAID clades (Fig 7).
196 The overall clades and lineages distribution highlighted the dominant occurrence of GRY

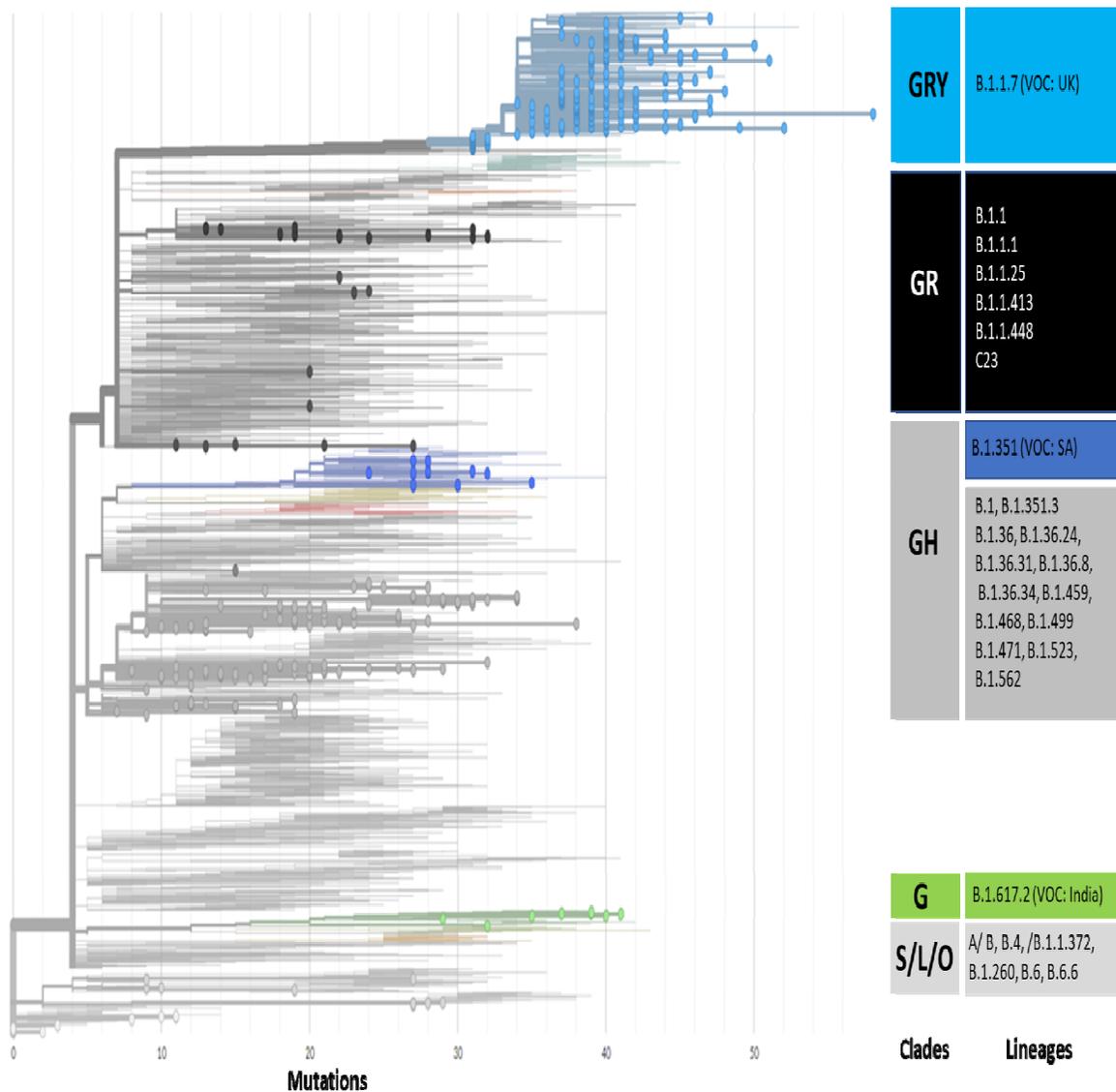
197 [B.1.1.7], GH [B.1, B.1.351, B.1.36, B.1.36.17, B.1.36.24, B.1.36.31, B.1.36.34, B.1.36.8,
198 B.1.468, B.1.471, B.1.499, B.1.523, B.1.562], GR [AE.4, B.1.1, B.1.1.1, B.1.1.25, B.1.1.372,
199 B.1.1.413, B.1.1.448, C.23], G [B.1.617.2], S [A], L [B, B.4] and O [B.1.260, B.6, B.6.6]
200 clades respectively. Maximum likelihood time resolved phylogeny tree of Pakistani and global
201 samples is presented in Fig 8. The genomic map of SARS-CoV-2 of Pakistani samples depicting
202 the mutational profiles is in given in Fig 6.

203



228
229
230 Fig 6. Genomic map of SARS-CoV-2 genomes of Pakistan. The reference genome in this
231 coverage map is Wuhan/Hu-1/2019.
232
233
234
235
236

237
238

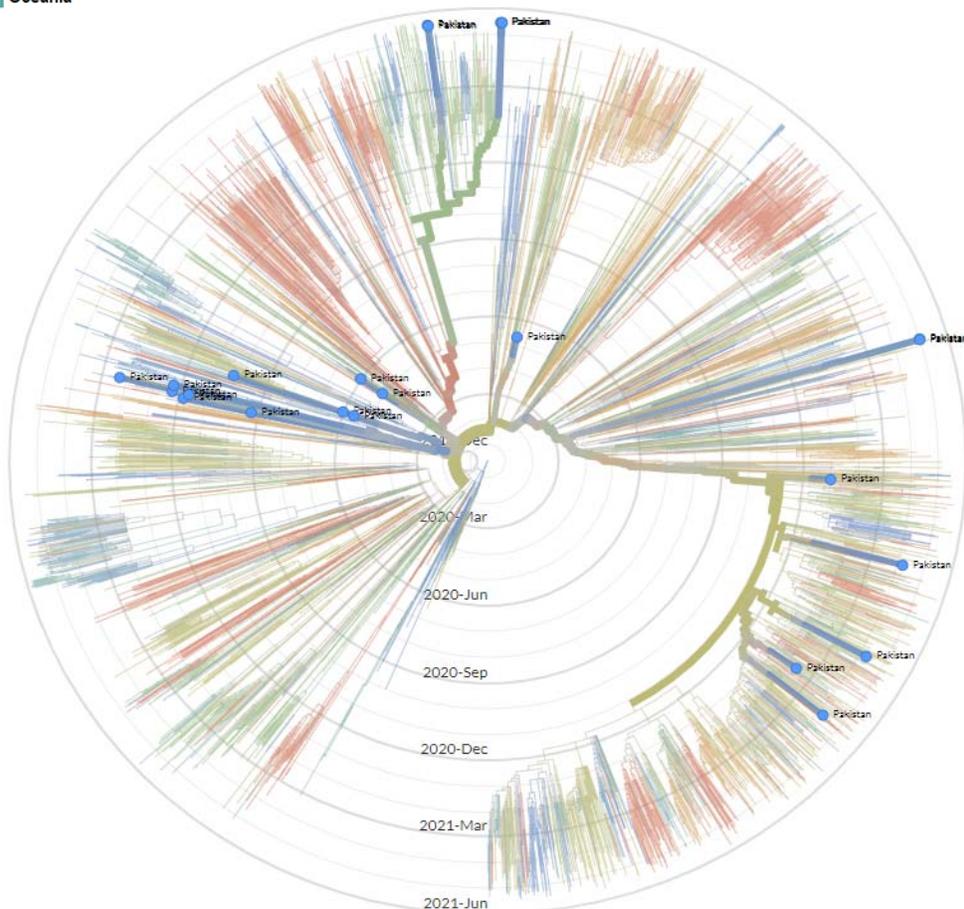


239
240
241
242
243
244
245
246
247
248

Fig 7. Maximum likelihood phylogenetic tree of Pakistani samples showing 7 clades and 29 lineages. The global SARS-CoV-2 genomes depicting the highest mutations ranging from 30 to 50 in Pakistani samples under the lineage of B.1.1.7, also known as UK variants of concern (UK-VOC).

249
250
251

- Asia
- Africa
- Europe
- South America
- North America
- Oceania



252
253 Fig 8. Maximum Likelihood time resolved phylogenetic tree of global samples representing the
254 emergence of SARS-CoV-2 in Pakistan.

255
256 **Wave-wise phylogenetic trees of SARS-CoV-2**

257 Maximum likelihood phylogenetic tree of first, second and third wave are presented in Fig. 9, 10
258 and 11. The tree of the first wave showed 16 lineages of 5 clades, with GH as major clade in this
259 wave (Fig. 9). During second wave 19 lineages belonging to four clades were observed and GH
260 was again major clade in this wave (Fig. 10). In third waves, 9 lineages representing 5 clades
261 were observed (Fig. 11), and GRY was the major clade in this wave (Fig. 12).

262

263
264

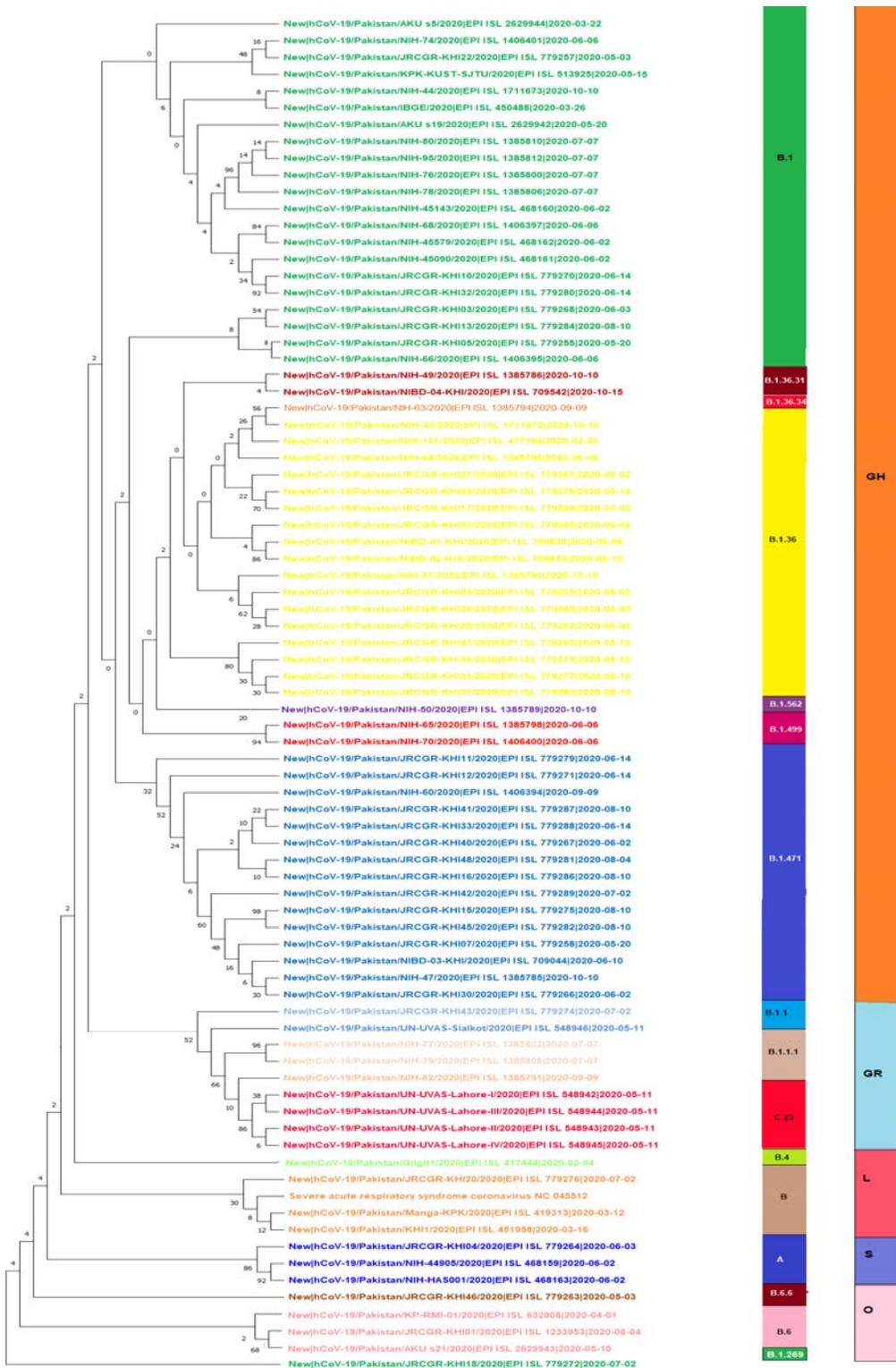
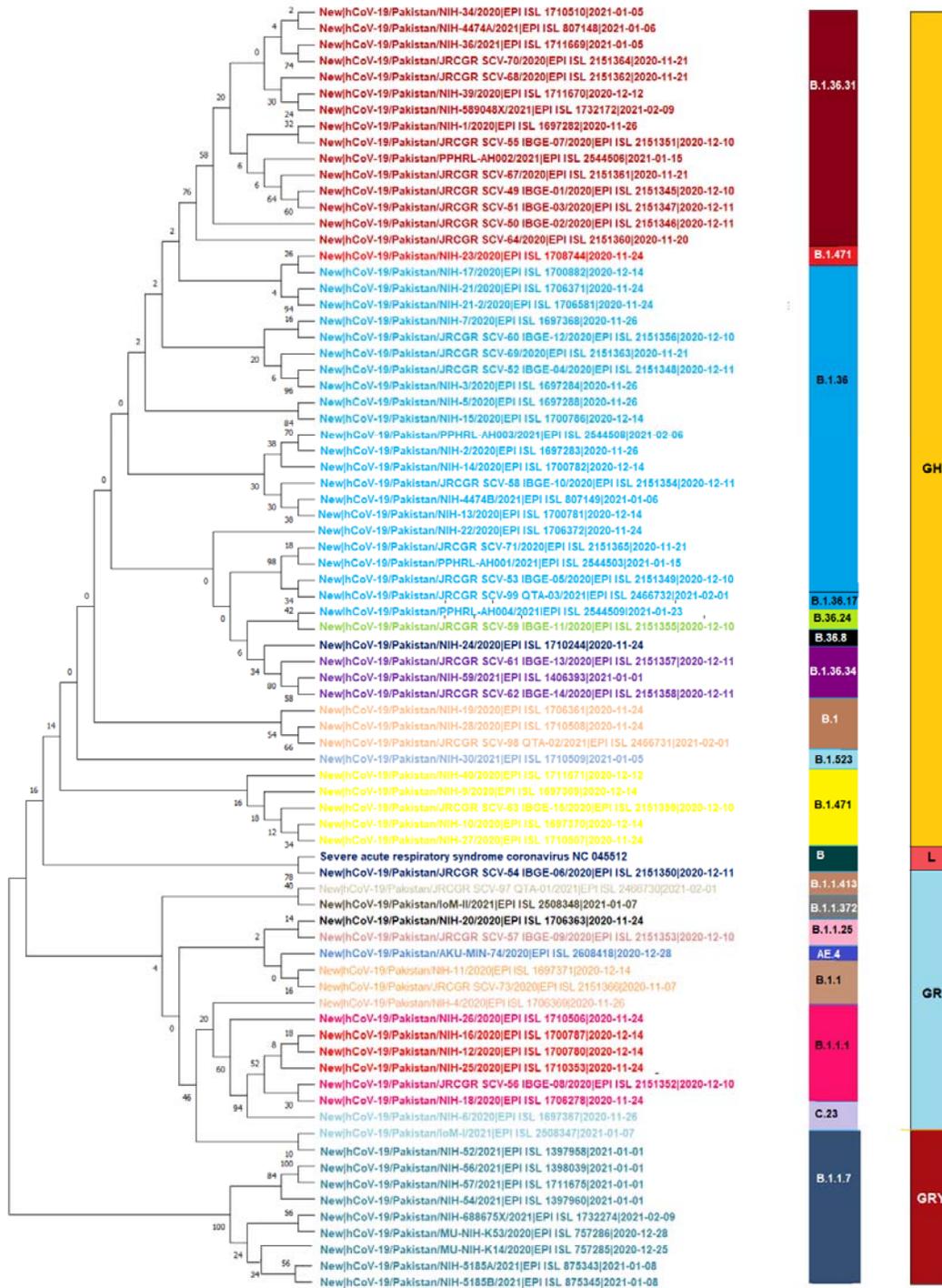
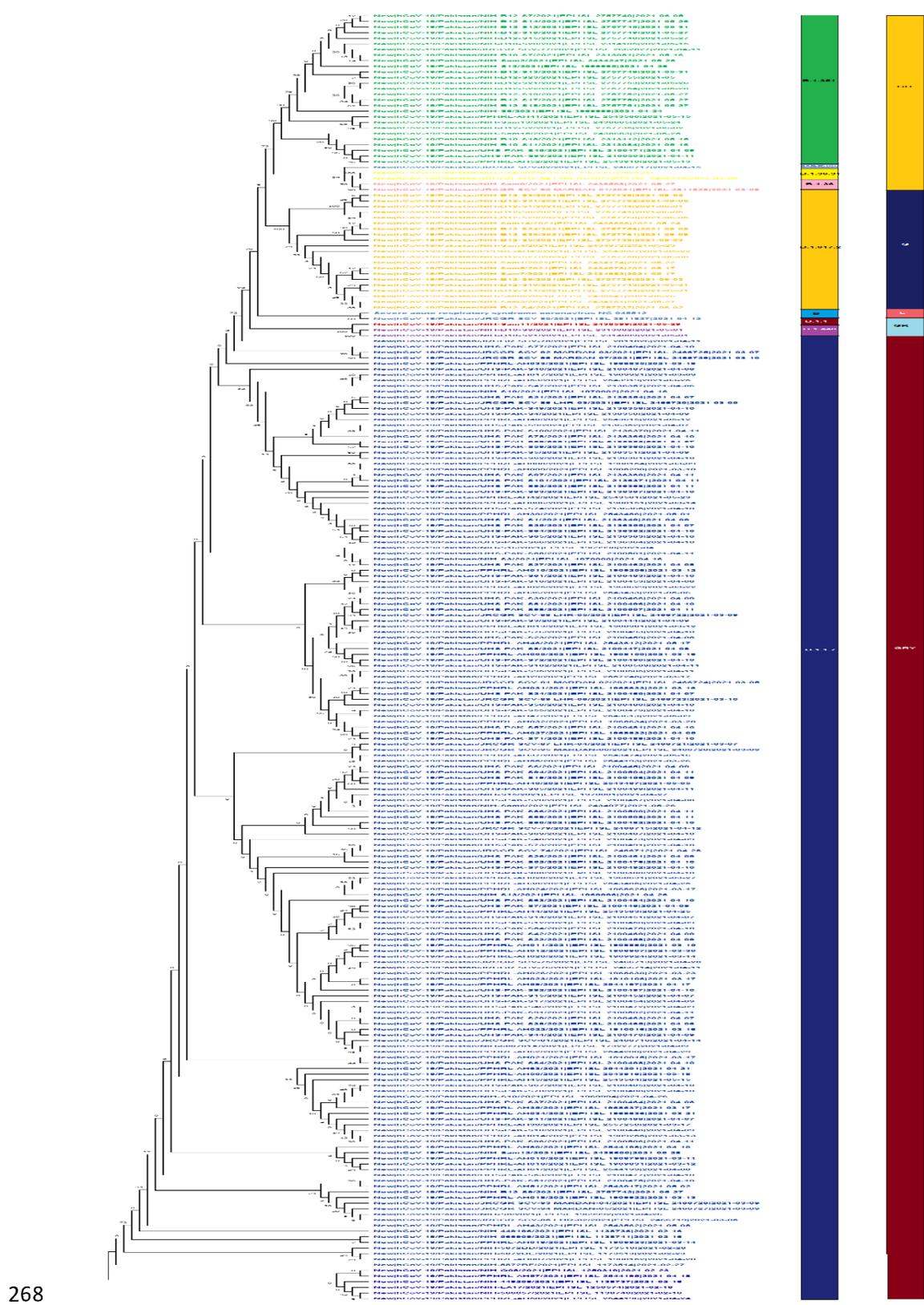


Fig 9. Maximum Likelihood phylogenetic tree of SARS-CoV-2 genome representing 16 lineages of 5 clades in first wave.



265

266 Fig 10. Maximum Likelihood phylogenetic tree of SARS-CoV2 genome representing 19 lineages
 267 and 4 clades in second wave.



268

269

270

Fig 11. Maximum Likelihood phylogenetic tree of SARS-CoV-2 genome representing 9 lineages and 5 clades in third wave.

271 **Discussion**

272 First two cases of COVID-19 were reported in Pakistan in February 2020 and since then country
273 has observed 978,662 infections and 22,642 deaths till 15 July 2021. However, the whole
274 genome sequences of first three SARS-CoV-2 samples, collected in March, revealed B and B.4
275 lineages. The lineage B.4 was reported to had major role in the early pandemic in Iran [22]. This
276 supports the hypothesis of an early transmission of virus to Pakistan most likely from Iran and
277 China, and this is also consistent with statement of Pakistan's health ministry officials who
278 confirmed first two cases of COVID-19 on 26th of February from patients travelled back to
279 Pakistan from Iran. It can be further confirmed by the fact that up to the first week of April 2020,
280 75% of the corona positive patients had the history of traveling back to Pakistan from Iran [23].
281 The occurrence of severe outbreaks of COVID-19 in two neighboring countries and declaration
282 of the disease as a pandemic by WHO prompted the Pakistan government to close the border
283 with China and put very strict screening methods at the Pak-Iran border [24, 25]. And by
284 following the guidelines of Federal government devised “*The National Action Plan for The*
285 *Corona Virus Disease (COVID-19) Pakistan*”, the quarantine centers were established in few big
286 cities including Islamabad [26, 27] and Pak-Iran border to help identify positive cases and
287 quarantine pilgrims returning back to Pakistan after spending time in Iran [28]. However, even
288 with all of these efforts, major lapses existed at every step mainly the uneven implementation of
289 immigrant policies to deal with the influx of people from the outside of the country [29], the lack
290 of facilities, poor infrastructure, and apathy & indifference among the peoples to follow
291 government devised Corona SOPs led to the spread of COVID-19 throughout the country [30].

292

293 **First wave of SARS-CoV-2**

294 Pakistan experienced three major waves of SARS-CoV-2 infection. First wave mainly covered
295 three months starting from May 2020 to July 2020 (Fig 2 & 3). From this wave, whole genome
296 sequences of 86 samples of SARS-CoV-2 were analyzed. These samples represented 16 lineages
297 belonging to 5 clades (Fig 9) including A(9.3%), B(4.6%), B.1(24.4%), B.1.1(2.3%),
298 B.1.1.1(3.4%), B.1.260(1.16%), B.1.36(18.6%), B.1.36.31(2.32%), B.1.36.34(1.2%),
299 B.1.471(18.6%), B.1.499(2.32%), B.1.562(1.2%), B.4(1.2%), B.6(3.4%), B.6.6(1.2%), and
300 C.23(4.65%). Lineage B.1 hit highest score in the first wave in Pakistan and internationally this
301 lineage was reported to have very high incidence in North America (the highest incidence), UK,
302 Germany, and Spain [31, 32]. It is also noteworthy that the highest proportion of Pakistani
303 diaspora is living in these countries. Therefore, it is highly likely that the transmission of this
304 lineage to Pakistan took place through international travelers most probably through the
305 travelling of Pakistani diaspora living in these countries. Though Pakistan suspended all
306 international flights from 21 March to 30 May 2020 but it is highly likely that before the
307 suspension of flight operations [33] many international passengers (positive for these lineage)
308 managed to enter the country. This hypothesis is further augmented from the fact that almost all
309 the sequence samples of this lineage were from big cities including Karachi, Capital city of
310 Islamabad, & Lahore, which have international airports. The lineage for second highest
311 incidence of infection of SARS-CoV-2 in Pakistan was B.1.36. This lineage had the highest
312 incidence of infections in India, Canada, UK and Hong Kong. Likewise, B.1.471 was also the
313 second most prevalent lineage with an incidence of 18.6%. The whole genome sequence data of
314 SARS-CoV-2 from GISAID predict the first incidence of this lineage in Saudi Arabia on 25th
315 March 2020 [34]. After that, this lineage was observed in Pakistan suggesting the direct
316 transmission of this lineage from Saudi Arabia to Pakistan. Likewise, the most likely reason for

317 the transmission of this lineage to Pakistan was again through the international travelers. Another
318 possible explanation for the transmission of these lineages could be the congregation (2,50,000
319 peoples) of Tablighi Jamaat in Pakistan in which in addition to Pakistan, 3000 delegates from 40
320 different countries did participate, with significant proportion of delegates from India [35]. This
321 hypothesis could be further confirmed from the timeline of the reports of these lineages in
322 Pakistan as the congregation took place in early March 2020 and these lineages were reported for
323 the first time in May and had the highest incidence in June 2020.

324 It is also worth mentioning that first 4 cases of C.23 lineage were reported for the first time
325 from Pakistan on 11-05-2020 and 3-4 week later this lineage was also reported from USA. It is
326 also highly likely that this lineage was transmitted to USA from Pakistan through the Pakistani
327 diaspora returning back to USA after visiting their homeland. However only 148 cases of this
328 lineage are reported globally up to 15 July 2021 indicating the lesser transmissibility of this
329 lineage.

330 **Second wave of SARS-CoV-2**

331 The second spell of SARS-CoV-2 infection in Pakistan ranged from the end of October 2020 to
332 mid of February 2021, however, it was at its peak from mid of November to mid of December
333 2020 (Fig 2 & 3). For this wave, after QC analysis 35 incomplete sequence of SARS-CoV-2
334 were removed and remaining 83 samples with whole genome sequences, available on GISAID,
335 were analyzed. According to our dataset, 18 lineages belonging to 4 clades (Fig. 10) were
336 observed in this wave out of which 10 were the same lineages which were already prevailing in
337 the country during the first wave and 8 others were the new lineages. The B.1.36 (initially
338 reported from India, Canada, and UK) which was the second most common (18.6%) lineage
339 during the first wave in Pakistan, had the highest incidence in the second wave as well. And

340 interestingly it was also observed that after the transmission of this lineage to Pakistan it
341 remained prevalent in the country and positive cases of this lineages were observed in almost
342 every month during the first and second wave. The B.1.36.31 (a sub-lineage of B.1.36) which
343 was also present in the first wave had the second highest incidence in second wave. The first case
344 of B.1.36.31 lineage was reported from Australia in the early March 2020 and then two cases of
345 this lineage were reported from England, and 4th case was reported from Pakistan. The incidence
346 of this lineage in the second wave remained 16.9% and out the total 532 genomic sequences, 18
347 were reported from the second wave in Pakistan (mostly from Karachi).

348 Likewise, the lineage B.1.471 which also had the second highest (18.6%) incidence during the
349 first wave, remained the third most common lineage in the second wave but its incidence was
350 reduced to 7.23%. It is also noteworthy that only 346 samples of this lineage were observed out
351 of global sequence data available at GISAID, out of which 22 were reported from Pakistan only.
352 And this lineage was also continuously observed from samples sequenced from May to
353 December 2020, during the first and second wave of SARS-CoV-2 infection, in the country.
354 However, B, B.1, B.1.1, and B.1.1.1, which were present in the first wave, got the incidences of
355 1.2%, 4.82%, 7.23%, & 8.4% respectively in the second wave as well.

356 Among the new lineages of second wave the B.1.1.7, B.1.36.34 and C.23 had 12.1%, 4.8%, and
357 2.4% incidences respectively. The first case of B.1.36.34 was reported from Canada in May,
358 2020 and (four month later) in September 2020 the genomic sequence of this variant was
359 reported from Pakistan [36]. It is most likely that both B.1.36.31, & B.1.36.34 were transmitted
360 to Pakistan through the overseas Pakistani, whose significant number live in Canada &
361 Australia and used to travel to Pakistan frequently. However, only 41 cases of B.1.36.34 are
362 reported globally, till 10 July 2021, out of which 5 were reported from Pakistan (Table 2). The

363 very small number of sequences of this lineage worldwide suggest its low transmissibility
364 compared with other lineages. Though, B.1.1.7 (UK-VOC) had the high incidence in the second
365 wave but it was reported almost at the end of the second wave and caused the highest infection
366 rate in the third wave.

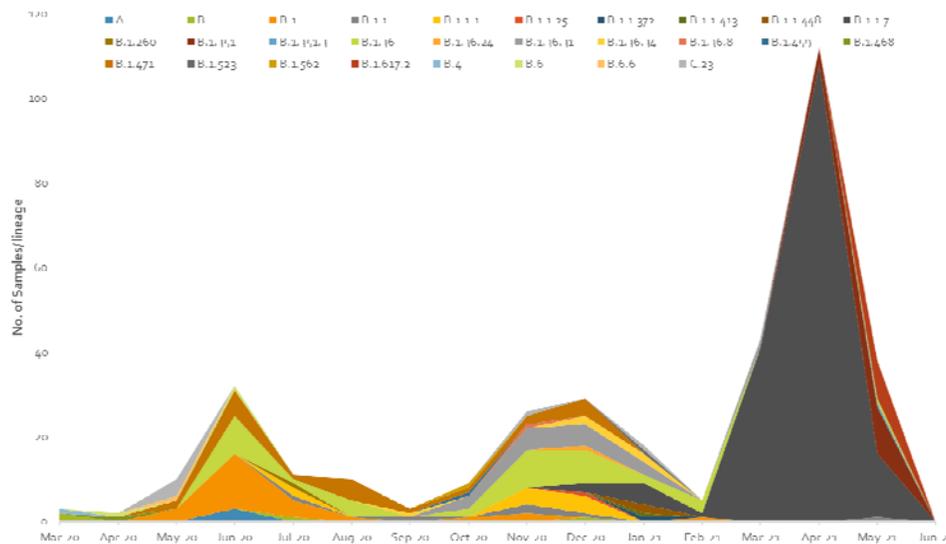
367 The B.1.36.17, B.1.36.24, B.1.1.25, B.1.1.372, B.1.1.413 were the minor lineages in second
368 wave and had their first cases reported mainly from England and Wales [36] and are also likely
369 to be transmitted to Pakistan from UK via Pakistani diaspora. Likewise, B.1.36.8, B.1.523, &
370 AE.4 had their first case reported from India, Saudi Arabia, and Bahrain respectively in the
371 early 2020 [37]. And they are also likely transmitted to Pakistan via international travelers.

372

373 **Third wave of SARS-CoV-2**

374 The third wave of SARS-CoV-2 proved worst in the country with the highest number of
375 infections (335,728) and deaths (7,849) compared with the previous waves. In the third wave,
376 B.1.1.7, B.1.617.2 and B.1.351 were the major lineages for the SARS-CoV-2 based infections
377 (Fig 11 & Fig 12). During this wave, first two cases of B.1.1.7 variants were reported on 20th
378 December 2020 in the country and kept spreading continuously and got the highest incidence
379 during the peak month of third wave. It is further evidenced by the submission of 108 (60.6%)
380 genomic sequences of B.1.1.7 (UK VOC), only in the April 2021, out of the total 178 genomic
381 sequence of this lineages from Pakistan, starting from 1st December 2020 to 10 July 2021 (Fig.
382 4). The B.1.1.7 variant is estimated to have emerged in September 2020 and quickly become the
383 dominant circulating variant in England[38]. In addition to Pakistan, this variant had been
384 reported from 30 other countries.

385 The modeled trajectory of this variant in the Pakistan revealed rapid increase in its cases in early
 386 2021, making it the predominant variant in the third wave (Fig 7, 8, & 9). Consistent, with our
 387 findings several other authors had reported that this variant had high transmissibility, virulence
 388 and death rates compared with other variants of SARS-CoV-2 [12, 39, 40]. Moreover, it is also
 389 worth mentioning that this variant had very high mutation rate (ranging from 45-50) (Fig. 7), in
 390 Pakistan, compared with other variants. Looking at the severity of the third wave, it is speculated
 391 that these mutations might have resulted in enhancing the transmissibility and pathogenicity of
 392 this variant in Pakistan.



393
 394 Fig 12. Pictorial view of genomic-data-based incidence of different lineages of SARS-CoV-2
 395 observed during the three waves of COVID-19 in Pakistan
 396

397 B.1.617.2 lineage (Delta variant) was another variant which was observed almost at the end
 398 (May 2021) third wave of COVID-19 in Pakistan. It was originally reported from India in April
 399 2020. Scientists have reported that Delta variant is 40-60% more contagious and virulent
 400 compared with B.1.1.7 [41, 42] and had caused a catastrophic wave of COVID-19 in India [43].
 401 However, Pakistan remained lucky enough that this lineage was contained very shortly and in

402 total only 24 genomic sequences of this variants had been reported from Pakistan, till 10-07-21.
403 However, only this variant is known to have caused at least 4 million deaths in the world, mainly
404 in India & UK [44].
405 B.1.351 lineage (Beta variant) was also reported during the third wave in Pakistan. Though this
406 variant is known as South African variant but its first few cases were reported from England in
407 the start of 2020 [37]. However, from Pakistan first case of this lineage was reported in May
408 2021 and it is highly likely that variant was transmitted to Pakistan from England through
409 overseas Pakistanis. This variant is also known to have significantly greater transmissibility over
410 alpha VOC [44-46]. However, only 30 genomic sequences of this variant are reported so far (10-
411 07-2021) from Pakistan which are 12.9% of total genome sequences of SARS-CoV-2 submitted
412 from Pakistan during the third wave of COVID-19. Our data revealed that B.1.1.7, B.1.351, and
413 B.1.617.2 were the dominant lineages in the third wave, and this is the most likely reason for the
414 severity of this wave in term of infections and deaths in the country.

415

416 **Materials and Methods**

417 **Collection of the health data of COVID-19 patients**

418 The data recorded on COVID-19 patients in different geographical regions of Pakistan as
419 Diseased, Hospitalized, Deceased and Recovered were taken from official website of Pakistan
420 (www.covid.gov.pk), for a period spanning from 3 January 2020 to 16 June 2021.

421

422 **Viral genome sequences**

423 In total, 453 whole genome sequences of SARS-CoV-2, reported from Pakistan, were
424 downloaded from GISAID (<https://www.gisaid.org/>) on 10 July 2021.

425 **Wave-wise categorization of SARS-CoV-2 genome sequences**

426 The data were categorized into three waves on the basis of date of collection of samples. For the
427 first wave which spanned from 1st January 2020 to 31 October 2020, 89 whole genome
428 sequences of SARS-CoV2 were available. Second wave ranged from 1st November 2020 to 15th
429 February 2021 and for this wave 118 samples of SARS-CoV-2 genome sequences were
430 submitted to GISAID. For third wave which spanned from Mid of February to the end of May
431 2021, a total of 246 whole genome sequences of SARS-CoV-2 were downloaded from the
432 GISAID.

433

434 **Multiple Sequence Alignment of SARS-CoV-2 genomes**

435 Sequence alignment of 453 Pakistani samples was performed using L-INS-I alignment method
436 implemented in MAFFT (v7.480), by setting data type as nucleic acids with gap extend penalty
437 of 0.123 and opening penalties default settings of 1.53 [47]. The sequences with more than 50
438 ambiguous bases were removed from the data file in order to minimize the number of false
439 positive. This resulted in the removal of 53 samples (3, 35, & 15 samples from 1st, 2nd & 3rd
440 wave respectively), and subsequently 400 samples were left for the analysis. The genome of
441 SARS-CoV-2 reported from Wuhan, China, in December 2019, and available at GenBank with
442 accession no. NC_045512.2 was used as a reference genome. Same method of alignment was
443 used when sequences of each wave were aligned separately.

444

445 **SARS-CoV-2 Lineage Assignment**

446 Phylogenetic Assignment of Named Global Outbreak Lineages (Pangolin) was used to describe

447 the genomic lineages of the Pakistani SARS-CoV-2 sequences. The Pangolin tool follows the
448 ‘Pango’ nomenclature system for classifying SARS-CoV-2 genomic sequences

449 **Construction of Phylogenetic tree with full length genomic sequences**

450 To infer our SARS-CoV-2 time-scaled and divergence phylogenetic tree, selected global
451 reference sequences were used for the Nextstrain analysis as of 10 July, 2021. This dataset
452 consisted of 3,804 sequences sampled between 26th December 2019 to 10 July 2021 from Africa
453 (541), Asia (937), Europe (727), North America (779), Oceania (512), and South America (566).
454 A collection of 400 Pakistani SARS-CoV-2 sequences was added to generate an initial dataset of
455 4204 whole-genome sequences. Using the Nextstrain metadata [21] to identify the accessions of
456 interest, the latest whole-genome sequence alignment from the GISAID database was
457 downloaded. These 4204 whole genome sequences were aligned using MAFFT (v.7.480) [47],
458 through multiple sequence alignment method, and manually edited by trimming the 5’ and 3’
459 untranslated regions and removing any gap only sites and low-quality sequence. The Wuhan/Hu-
460 1/2019 (EPI_ISL_402125), sampled on December 31, 2019, from Wuhan, China was
461 downloaded from the GISAID and used as reference genome. Finally, the maximum likelihood
462 phylogenetic tree was built by using the Nextstrain pipelines which incorporates Augur for
463 generation of phylogenetic tree and Auspice for visualizations. A coverage map was created
464 using Nextclade (version 0.7.5), which performs banded Smith–Waterman alignment with an
465 affine gap penalty.

466 **Wave-wise construction of phylogenetic tree**

467 Phylogenetic tree of each wave was inferred by using the Maximum Likelihood method based on
468 amino acid substitution model of Jones Taylor Thornton (JTT) matrix-based model in Mega-X
469 software. In each wave, initial phylogenetic tree(s) for the heuristic search were obtained

470 automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances
471 estimated using the JTT model, and then finally, maximum likelihood phylogenetic tree of each
472 wave was made by selecting the topology with superior log likelihood value.

473 **Acknowledgement**

474 We are thankful to all the laboratories that submitted their sequences to the GISAID platforms.

475 We are also grateful to GISAID team for managing this platform.

476

477

478 **Conflicts of Interest**

479 The authors declare that they have no conflict of interest. This study had no role in the patient
480 consent under the current frame of research.

481

482 Reference

- 483 1. World Health Organization. Director-General's remarks at the media briefing on 2019-nCoV on
484 11 February 2020. [http://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-](http://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020)
485 [media-briefing-on-2019-ncov-on-11-february-2020](http://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020) (Accessed on February 12, 2020).
- 486 2. Centers for Disease Control and Prevention. 2019 Novel coronavirus, Wuhan, China. Information
487 for Healthcare Professionals. <https://www.cdc.gov/coronavirus/2019-nCoV/hcp/index.html> (Accessed
488 on June 28, 2021).
- 489 .
- 490 3. Alteri C, Cento V, Piralla A, Costabile V, Tallarita M, Colagrossi L, et al. Genomic epidemiology of
491 SARS-CoV-2 reveals multiple lineages and early spread of SARS-CoV-2 infections in Lombardy, Italy.
492 *Nature Communications*. 2021;12(1):434. doi: 10.1038/s41467-020-20688-x.
- 493 4. Alouane T, Laamarti M, Essabbar A, Hakmi M, Bouricha EM, Chemao-Elfihri MW, et al. Genomic
494 Diversity and Hotspot Mutations in 30,983 SARS-CoV-2 Genomes: Moving Toward a Universal Vaccine
495 for the "Confined Virus"? *Pathogens* (Basel, Switzerland). 2020;9(10). Epub 2020/10/15. doi:
496 10.3390/pathogens9100829. PubMed PMID: 33050463; PubMed Central PMCID: PMCPMC7600297.
- 497 5. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019
498 novel coronavirus: implications for virus origins and receptor binding. *Lancet* (London, England).
499 2020;395(10224):565-74. Epub 2020/01/30. doi: 10.1016/S0140-6736(20)30251-8. PubMed PMID:
500 32007145.
- 501 6. Smith EC, Sexton NR, Denison MR. Thinking Outside the Triangle: Replication Fidelity of the
502 Largest RNA Viruses. *Annual Review of Virology*. 2014;1(1):111-32. doi: 10.1146/annurev-virology-
503 031413-085507. PubMed PMID: 26958717.
- 504 7. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. Coronaviruses. *RNA Biology*.
505 2011;8(2):270-9. doi: 10.4161/rna.8.2.15013.
- 506 8. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature
507 proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*.
508 2020;5(11):1403-7. doi: 10.1038/s41564-020-0770-5.
- 509 9. European Centre for Disease Prevention and Control. Rapid increase of a SARS-CoV-2 variant
510 with multiple spike protein mutations observed in the United Kingdom, December 2020.
511 [https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-](https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf)
512 [mutations-United-Kingdom.pdf](https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf) (Accessed on December 21, 2020).
- 513 10. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing transmissibility of
514 SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. 2021;593(7858):266-9. doi: 10.1038/s41586-021-03470-
515 x.
- 516 11. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L. Risk of mortality
517 in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ*.
518 2021;372:n579-n. doi: 10.1136/bmj.n579. PubMed PMID: 33687922.
- 519 12. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated
520 transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* (New York, NY).
521 2021;372(6538):eabg3055. Epub 2021/03/03. doi: 10.1126/science.abg3055. PubMed PMID: 33658326.
- 522 13. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a
523 SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021;592(7854):438-43. doi: 10.1038/s41586-
524 021-03402-9.
- 525 14. Galloway SE, Paul P, MacCannell DR, Johansson MA, Brooks JT, MacNeil A, et al. Emergence of
526 SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021. *MMWR Morbidity*
527 *and mortality weekly report*. 2021;70(3):95-9. doi: 10.15585/mmwr.mm7003e2. PubMed PMID:
528 33476315.

- 529 15. Yang W, Shaman J. Epidemiological characteristics of three SARS-CoV-2 variants of concern and
530 implications for future COVID-19 pandemic outcomes. Cold Spring Harbor Laboratory; 2021.
- 531 16. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Increased Resistance of SARS-CoV-2
532 Variants B.1.351 and B.1.1.7 to Antibody Neutralization. bioRxiv : the preprint server for biology.
533 2021:2021.01.25.428137. doi: 10.1101/2021.01.25.428137. PubMed PMID: 33532778.
- 534 17. Dejnirattisai W, Zhou D, Supasa P, Liu C, Mentzer AJ, Ginn HM, et al. Antibody evasion by the P.1
535 strain of SARS-CoV-2. Cell. 2021;184(11):2939-54.e9. Epub 2021/03/30. doi: 10.1016/j.cell.2021.03.055.
536 PubMed PMID: 33852911.
- 537 18. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, et al. Genomics and
538 epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science. 2021;372(6544):815-21. Epub
539 2021/04/16. doi: 10.1126/science.abh2644. PubMed PMID: 33853970; PubMed Central PMCID:
540 PMC8139423.
- 541 19. Zhang W, Davis BD, Chen SS, Martinez JMS, Plummer JT, Vail E. Emergence of a novel SARS-CoV-
542 2 strain in Southern California, USA. JAMMA. 2021;325(13). doi: 10.1101/2021.01.18.21249786.
- 543 20. Deng X, Garcia-Knight MA, Khalid MM, Servellita V, Wang C, Morris MK, et al. Transmission,
544 infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R
545 spike protein mutation. medRxiv. 2021:2021.03.07.21252647. doi: 10.1101/2021.03.07.21252647.
546 PubMed PMID: 33758899.
- 547 21. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time
548 tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121-3. doi: 10.1093/bioinformatics/bty407
549 %J Bioinformatics.
- 550 22. Zohreh Fattahi MM, Khadijeh Jalalvand, Fatemeh Aghakhani Moghadam, Azam
551 Ghaziasadi, Fatemeh Keshavarzi, Jila Yavarian, Ali Jafarpour, Seyedeh Elham Mortazavi, Fatemeh
552 Ghodratpour, Hanieh Behravan, Mohammad Khazeni, Seyed Amir Momeni, Issa Jahanzad, Abdolvahab
553 Moradi, Alijan Tabarraei, Sadegh Ali Azimi, Ebrahim Kord, Seyed Mohammad Hashemi-Shahri, Azarakhsh
554 Azaran, Farid Yousefi, Zakiye Mokhames, Alireza Soleimani, Shokouh Ghafari, Masood Ziaee, Shahram
555 Habibzadeh, Farhad Jeddi, Azar Hadadi, Alireza Abdollahi, Gholam Abbas Kaydani, Saber Soltani, Talat
556 Mokhtari-Azad, Reza Najafipour, Reza Malekzadeh, Kimia Kahrizi, Seyed Mohammad Jazayeri, Hossein
557 Najmabadi. SARS-CoV-2 outbreak in Iran: the dynamics of the epidemic and evidence on two
558 independent introductions. Transboundary and Emerging Diseases.
559 2021; <https://doi.org/10.1111/tbed.14104>.
- 560 23. Jamil B. Clinical features, diagnosis and management of COVID-19 patients in the outdoor
561 setting. Clinical Medicine. 2020;70(5).
- 562 24. National Action Plan for Corona Virus Disease (COVID-19) Pakistan. Available online at:
563 <https://www.nih.org.pk/wp-content/uploads/2020/03/COVID-19-NAP-V2-13-March-2020.pdf>.
564 Islamabad2020.
- 565 25. Khan Z, Muhammad K, Ahmed A, Rahman H. Coronavirus outbreaks: prevention and
566 management recommendations. Drugs Ther Perspect. 2020:1-3. doi: 10.1007/s40267-020-00717-x.
567 PubMed PMID: 32218651.
- 568 26. Coronavirus patients, suspects break out of Sukkur quarantine facility. ARY News. (2020).
569 Available online at: <https://arynews.tv/en/sukkur-quarantinefacility-coronavirus-suspects/>.
- 570 27. 1,200-bed quarantine facility inaugurated at Karachi Expo Centre. The Nation. (2020). Available
571 online at: [https://nation.com.pk/02-Apr-2020/1-200-bedquarantine-facility-inaugurated-at-karachi-
572 expo-centre](https://nation.com.pk/02-Apr-2020/1-200-bedquarantine-facility-inaugurated-at-karachi-expo-centre).
- 573 28. Pakistan's Spike in Coronavirus Cases Raises Quarantine Concerns. Al Jazeera (2020). Available
574 online at: [https://www.aljazeera.com/news/2020/03/pakistan-spike-coronavirus-cases-raises-
575 quarantine-concerns-200316090946936.html](https://www.aljazeera.com/news/2020/03/pakistan-spike-coronavirus-cases-raises-quarantine-concerns-200316090946936.html).

- 576 29. Coronavirus: Travel Restrictions, Border Shutdowns by Country. Al Jazeera (2020). Available
577 online at: <https://www.aljazeera.com/news/2020/03/coronavirus-travel-restrictions-border-shutdowns->
578 [country-200318091505922.html](https://www.aljazeera.com/news/2020/03/coronavirus-travel-restrictions-border-shutdowns-country-200318091505922.html).
- 579 30. Dealing with Covid-19 in Pakistan. News. (2020). Available online at:
580 <https://www.thenews.com.pk/print/629602-dealingwith-covid-19-in-pakistan>.
- 581 31. Stange M, Mari A, Roloff T, Seth-Smith HM, Schweitzer M, Brunner M, et al. SARS-CoV-2
582 outbreak in a tri-national urban area is dominated by a B.1 lineage variant linked to a mass gathering
583 event. PLoS Pathog. 2021;17(3). doi: 10.1371/journal.ppat.1009374.
- 584 32. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, et al. The emergence of SARS-CoV-2 in
585 Europe and the US. bioRxiv : the preprint server for biology. 2020:2020.05.21.109322. doi:
586 10.1101/2020.05.21.109322. PubMed PMID: 32511416.
- 587 33. The Guardian. Pakistan suspends international flights for two weeks. Available from
588 [https://www.theguardian.com/world/2020/mar/21/pakistan-suspends-international-flights-for-](https://www.theguardian.com/world/2020/mar/21/pakistan-suspends-international-flights-for-two-weeks-to-stem-covid-19-spread-2020)
589 [two-weeks-to-stem-covid-19-spread 2020](https://www.theguardian.com/world/2020/mar/21/pakistan-suspends-international-flights-for-two-weeks-to-stem-covid-19-spread-2020) [cited 2020 13 July].
- 590 34. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality.
591 Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease
592 bulletin. 2017;22(13):30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494. PubMed PMID: 28382917.
- 593 35. Tableeghi Jamaat in hot water in Pakistan too for Covid-19 spread (Accessed on 20 June 2021).
594 Available on <https://www.dawn.com/news/1547354> 2020.
- 595 36. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality.
596 Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease
597 bulletin. 2017;22(13). Epub 2017/04/07. doi: 10.2807/1560-7917.Es.2017.22.13.30494. PubMed PMID:
598 28382917; PubMed Central PMCID: PMC5388101.
- 599 37. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to
600 global health. Global Challenges. 2017;1(1):33-46. doi: <https://doi.org/10.1002/gch2.1018>.
- 601 38. Public Health England. Investigation of novel SARS-CoV-2 variant: variant of concern 202012/01,
602 technical briefing 3. London, United Kingdom: Public Health England; 2020.
603 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/95](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950823/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3_-_England)
604 [0823/Variant of Concern VOC 202012 01 Technical Briefing 3 - England](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950823/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3_-_England).
- 605 39. Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, et al. Genomic characteristics and
606 clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing
607 and hospital-based cohort study. Lancet Infect Dis. 2021:S1473-3099(21)00170-5. doi: 10.1016/S1473-
608 3099(21)00170-5. PubMed PMID: 33857406.
- 609 40. Haynes WA, Kamath K, Lucas C, Shon J, Iwasaki A. Impact of B.1.1.7 variant mutations on
610 antibody recognition of linear SARS-CoV-2 epitopes. Cold Spring Harbor Laboratory; 2021.
- 611 41. Yadav PD, Sapkal GN, Abraham P, Ella R, Deshpande G, Patil DY, et al. Neutralization of variant
612 under investigation B.1.617 with sera of BBV152 vaccinees. Cold Spring Harbor Laboratory; 2021.
- 613 42. Hoffmann M, Hofmann-Winkler H, Krüger N, Kempf A, Nehlmeier I, Graichen L, et al. SARS-CoV-2
614 variant B.1.617 is resistant to Bamlanivimab and evades antibodies induced by infection and vaccination.
615 Cold Spring Harbor Laboratory; 2021.
- 616 43. Tada T, Zhou H, Dcosta BM, Samanovic MI, Mulligan MJ, Landau NR. The Spike Proteins of SARS-
617 CoV-2 B.1.617 and B.1.618 Variants Identified in India Provide Partial Resistance to Vaccine-elicited and
618 Therapeutic Monoclonal Antibodies. Cold Spring Harbor Laboratory; 2021.
- 619 44. Williams H, Hutchinson D, Stone H. Watching Brief: The evolution and impact of COVID-19
620 variants B.1.1.7, B.1.351, P.1 and B.1.617. Global Biosecurity. 2021;3(1). doi: 10.31646/gbio.112.
- 621 45. Roquebert B, Trombert-Paolantoni S, Haim-Boukobza S, Lecorche E, Verdurme L, Foulongne V,
622 et al. The SARS-CoV-2 B.1.351 lineage (VOC β) is outgrowing the B.1.1.7 lineage (VOC α) in some French
623 regions in April 2021. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European

624 communicable disease bulletin. 2021;26(23):2100447. doi: 10.2807/1560-7917.ES.2021.26.23.2100447.
625 PubMed PMID: 34114541.
626 46. Shen X, Tang H, Pajon R, Smith G, Glenn GM, Shi W, et al. Neutralization of SARS-CoV-2 Variants
627 B.1.429 and B.1.351. The New England journal of medicine. 2021;384(24):2352-4. Epub 2021/04/07. doi:
628 10.1056/NEJMc2103740. PubMed PMID: 33826819.
629 47. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence
630 and structural alignment. Nucleic Acids Research. 2019;47(W1):W5-W10. doi: 10.1093/nar/gkz342 %J
631 Nucleic Acids Research.
632