



Perspective

Footprints of SARS-CoV-2 genome diversity in Pakistan, 2020–2021

Zaira Rehman^{a,*}, Massab Umair^{a,*}, Aamer Ikram^a, Aammad Fahim^b, Muhammad Salman^a^a National Institute of Health (NIH), Islamabad, 44000, Pakistan^b Maroof International Hospital, Islamabad, 44000, Pakistan

The rapid spread of SARS-CoV-2 has significantly impacted the worldwide health system. The SARS-CoV-2 currently bears a remarkably low genetic diversity even though it carries one of the largest RNA genomes among viruses (Rausch et al., 2020). However, the coronaviruses harbor the capability of undergoing recombination at a high rate which can lead to the emergence of novel viral derivatives (Rausch et al., 2020; Gribble et al., 2021). This in turn requires not only global surveillance of SARS-CoV-2 genome in various countries but also careful scrutiny in animal genomic reservoirs. Conventionally, RNA viruses evolve with a high mutation rate, however, the presence of ExoN ribonuclease in SARS-CoV-2 genome has made its case different from other viral species (Gribble et al., 2021). The variables of natural selection which potentially drift the SARS-CoV-2 evolutionary dynamics can be recorded by analyzing deposited sequence genomes for its fitness, transmissibility potential, and pathogenicity (Rouchka et al., 2020). This can potentially provide a way to draw a holistic picture at a national level, while simultaneously providing a comparative overview with worldwide sequences.

The first confirmed case of SARS-CoV-2 was reported on February 26, 2020 in Pakistan. Since then the number of cases started to increase gradually and peaked in the mid of June 2020, with greater than 6,000 confirmed cases per day (Government of Pakistan, 2021). The number of cases then started to decline in August 2020 and produced a flattened curve during August and September 2020. The government announcement of the second wave hitting Pakistan was made on October 28, 2020 with positivity rate between 18% and 23%, which lasted for three months and ended in January 2021 (Ali, 2020). The third wave of outbreak struck Pakistan after the identification of B.1.1.7 SARS-CoV-2 variant in December 2020. The number of cases steadily arose particularly catching their pace in the first week of March 2021 and peaked in late April 2021 with the COVID-19 cases averaging more than 5,500 cases per day, very near to the maximum average recorded in June 2020 (Government of Pakistan, 2021). The decline in number of positive cases during third wave was observed in June 2021 that again followed a sharp rise in start of July 2021, marking the commencement of fourth wave. The fourth wave kept its presence in Pakistan for roughly three months

and declined in October 2021, that led to a flattened curve in November and December 2021. Pakistan witnessed its first case of Omicron variant being reported on December 13, 2021. Since then the COVID-19 cases particularly attributable to Omicron variant, are on the rise, and till December 27, 2021, the cases of Omicron in Pakistan have been 75. The state health authorities has warned this resurgence in COVID-19 cases as a marker for fifth wave on January 3, 2022. Up to December 2021, a total of 1,745 whole-genome sequences of SARS-CoV-2 has been reported on GISAID and NCBI from Pakistan.

The PANGOLIN database (Rambaut et al., 2020) was used for the analysis of SARS-CoV-2 lineages circulating in Pakistan. During the first year (i.e. year 2020) of the pandemic only a small number of sequences (n = 296) were reported from Pakistan. But in 2021, the sequencing capacity increased and the number of sequences reported has increased and reached up to 1,745 whole genomes from Pakistan. The analysis indicated that about 58% of isolates were classified as B.1 from April to June 2020 (the period of the first wave). The next lineages along with B.1 were B.1.1.1, B.1.36 and A. The lineage A and B.1.1.1 were not observed in the isolates sequenced after August 2020. The sporadic cases of B.1 and B.1.36 were observed after August 2020. During the second wave, different lineages (B1.36.31, B1.471, and B1.562) were observed with a large number of isolates belonging to B.1.36.31. The sequences reported during December 2020 and January 2021 were related to B.1.1.7 (Alpha) lineage (Fig. 1A). The start of the third wave and peak after March was confirmed by the presence of B.1.1.7 lineage. Other sub-lineages of B.1.1.7 such as Q.1 and Q.4 have also shown up sporadically from April 2021 but they are not consistently visible in a significant number of sequences in successive months. Interestingly, B.1.36.31 lineage was not found in any of sequences studied in successive months after April 2021. The B.1.351 (Beta) first observed in Pakistan in March 2021 and highest number of cases was reported during May–June 2021. However, the lineage B1.617.2 (Delta) first showed its presence in April 2021, and its number has been steadily increasing with a more than 100% increase in number in August 2021 compared to sequences from July 2021. During May–June 2021 (the time of third wave), the major lineages were found to be the B.1.1.7, B.1.351 and B.1.617.2. The B.1.617.2 and its

* Corresponding authors. National Institute of Health (NIH), Islamabad 44000, Pakistan.

E-mail addresses: rehman.zaira@yahoo.com (Z. Rehman), m.umair@nih.org.pk (M. Umair).

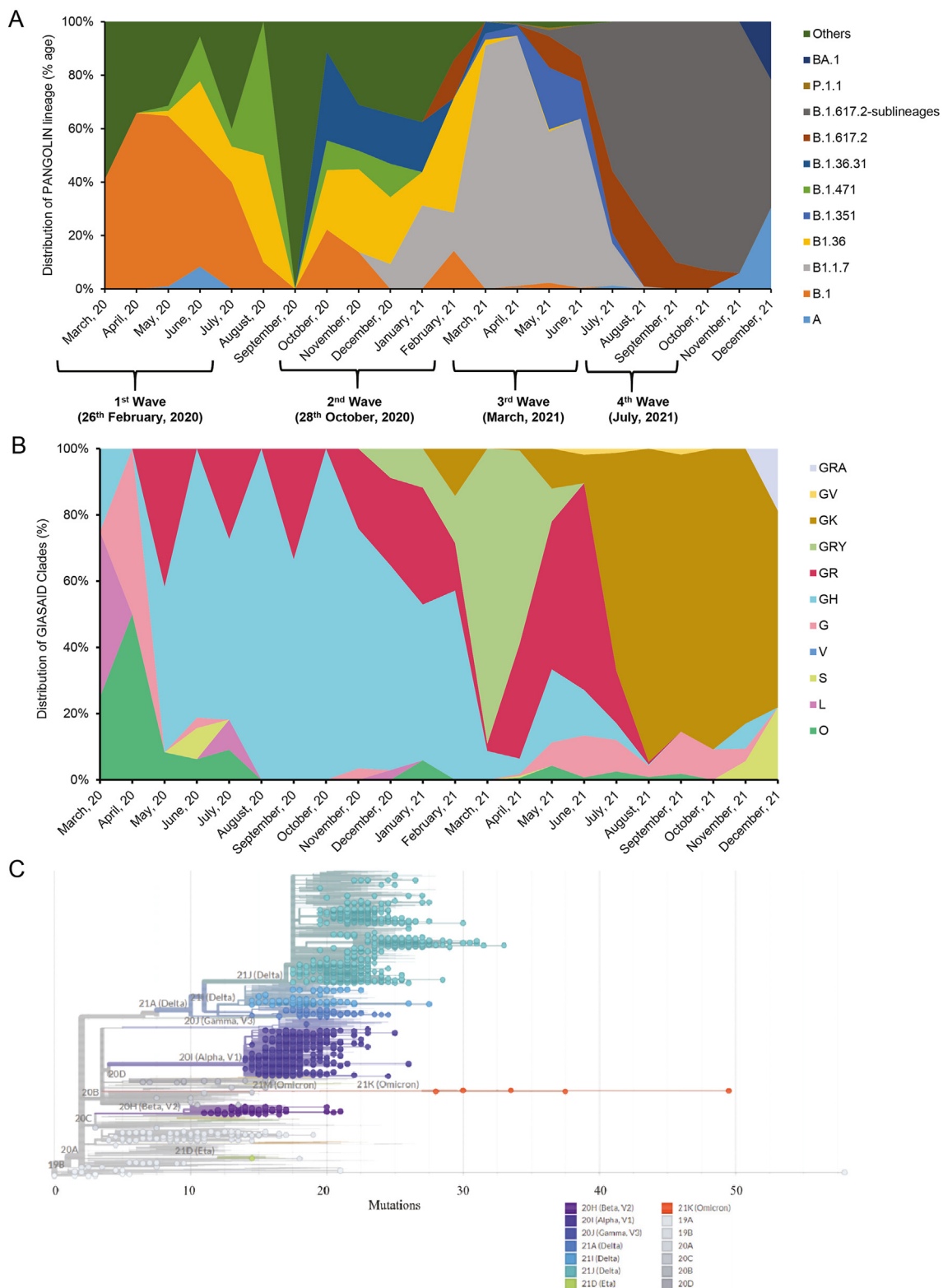


Fig. 1. Distribution of clades and lineages of total SARS-CoV-2 genomes from Pakistan. **A** PANGOLIN lineages of the SARS-CoV-2 isolates circulating in Pakistan. The majority of isolates were classified as B.1.617.2 and B.1.1.7 lineage. The B.1.617.2-sublineages include: AY.1, 0.102, 0.103, 0.108, 0.11, 0.111, 0.112, 0.114, 0.116, 0.118-0.122, 0.125-0.127, 0.13, 0.20, 0.33, 0.36, 0.39, 0.4, 0.42-0.45, 0.46, 0.46.6, 0.5.4, 0.51, 0.55, 0.56, 0.58, 0.59, 0.61, 0.62, 0.64, 0.65, 0.7.1, 0.70, 0.73, 0.75, 0.77, 0.79, 0.84, 0.86, 0.88, 0.9.2, 0.92, 0.96, 0.98.1, 0.99. Others include: AE.4, B, B.4, B.6, B.1.1, B.1.1.1, B.1.319, B.1.260, B.1.564, B.1.488, B.1.1.372, B.1.36.17, B.1.523, C.23, B.1.1.413, B.1.356, B.1.468, A.27, B.1.525, C.23, B.1.562, B.1.36.8, B.1.36.20, B.1.36.24, and B.1.36.34. **B** Distribution of GISAID clades of SARS-CoV-2 genomes from Pakistan. The majority of isolates belong to GK, GR, and GH clade. **C** Maximum likelihood phylogeny of 1745 SARS-CoV-2 genomes from Pakistan. Branches with circles represent the genomes analyzed in the current study.

sub-lineages are the most prevalent lineages in our sequenced samples from the months of August–October 2021 (the time of fourth wave). There were 52 sub-lineages of B.1.617.2 observed in studied isolates. The sub-lineages first appeared in May 2021 and the highest number of sub lineages were identified in the month of August 2021. On 13th December 2021, a new lineage BA.1 (Omicron) has been identified in Pakistan. Interestingly, along with the BA.1, the lineage A has also been observed in the isolates.

The overall distribution of GISAID clades over time of all the reported sequences is shown in Fig. 1B. During the period from March to July 2020, the sequences from all the clades (O/L/S/G/GH/GR) were observed in Pakistan with the most dominant clade GH. In September and October 2020, only the GH clade was observed. The sequences from December 2020 and January 2021 belonged to GRY clade reaching their maximum number in April 2021. However, it has not been visible after May 2021. The GK clade had first shown its appearance in February 2021 and became the most dominant clade in the sequences reported from July to December 2021, followed by the GR clade which showed its peak in June 2021. No sequence of GR clade was observed after August 2021. The GV clade was only observed in the sequences only from June, July, and September 2021. However, it hasn't been observed in sequences after September 2021. In December 2021 a new GRA clade has shown its emergence. Interestingly, the S clade has also re-emerged in November 2021 with number of cases rising in December 2021.

Phylogenetic analysis of domestic strains using NextStrain (Hadfield et al., 2018) revealed the presence of 20A clade during the first wave of COVID-19. Other major clades identified during this period were 19A and 20D (Fig. 1C). The isolates from the second wave belonged to 20A clade. During the third and fourth wave, the major clades were 20H, 20I, 21J, and 21I. The 21K clade was shown its appearance in December 2021.

Since the first reporting of the SARS-CoV-2 genome sequence on January 10, 2020, there has been an explosion of SARS-CoV-2 genome sequence deposits on GISAID numbering to 6,510,899 whole genome sequences as of December 31, 2021. These sequence deposits reflect the active effort of the worldwide health agencies across the globe in genomic surveillance of SARS-CoV-2 for identification of new mutant strains which can potentially alter the course of national public health programs.

The data points related to first and second wave of pandemic showed a rather less heterogeneous genomic diversification of SARS-CoV-2 in Pakistan. That's because the COVID-19 samples cohort from first and second wave probably have not been harboring new mutant strains. However, the third wave not only brought new mutant strains of SARS-CoV-2 but with rather differing clade distribution. During the third wave of pandemic, B.1.1.7, B.1.351, and B.1.617.2 lineages has been observed, while in the fourth wave, B.1.617.2 and its sub lineages have a major share. The emergence of alpha variant within Pakistani population has been formally reported earlier this year and the delta variant in April 2021.

Pakistan is currently experiencing the commencement of fifth wave of SARS-CoV-2 infectivity, with the proven emergence of Omicron variant from both the northern and southern parts. Keeping in view of the previous resurgence pattern of COVID-19 cases of each pandemic wave, with the third and fourth wave being relatively larger than the second wave (Government of Pakistan, 2021), the likelihood of other strains

circulating in Pakistani population is not completely unlikely. The emergence of new mutant strains in cases of introductions from international travelers and local festivals has been observed in this geographical region previously. The emergence of double mutant strain in India (Science Media Center, 2021), while there was a religious festival going on the borders of Punjab Province with the arrival of pilgrimages in April 2021 (Daily Sabah, 2021), followed by the appearance of delta variant in Pakistani population, further point towards the likelihood of dissemination of new variants from travelers. To add further, previously, the importation of SARS-CoV-2 particularly from the expatriate Pakistani travelers from UK, USA, Saudi Arabia and pilgrimage travels from Iran have contributed to the SARS-CoV-2 genomic diversity in the Pakistani population. This demands more effective and vigilant genomic surveillance of SARS-CoV-2 strains especially in the hotspot areas.

Furthermore, with the COVID-19 vaccination coverage programs across the world, vaccine escape mutants of SARS-CoV-2 have also emerged in various parts which reiterate the need for watchful surveillance and its potential emergence in Pakistan. In Pakistan, the vaccination rate is increasing with roughly 30% of the population being fully vaccinated. Despite the vaccination in countries like France and Italy, the number of COVID19 cases has increasing rapidly due to emergence of Omicron variant, particularly considering its potential in infecting already vaccinated individuals. The lesson from worldwide genomic surveillance of SARS-CoV-2 is fully applicable in Pakistan as well where flattening the pandemic curve and continued improvement of public health programs can help in keeping the lid on mortality and morbidity numbers of the COVID-19 pandemic.

Footnotes

No funding was acquired for the study. The authors declare they have no conflict of interest.

References

- Ali, A., 2020. Second wave of COVID-19 in Pakistan: Are more episodes down the road? <https://www.bmj.com/content/371/bmj.m4113/rr-0>. (Accessed 21 April 2021).
- Daily Sabah, 2021. Around 1,000 Indian Sikh pilgrims arrive in Pakistan. <https://www.dailysabah.com/world/asia-pacific/around-1000-indian-sikh-pilgrims-arrive-in-pakistan>. (Accessed 21 April 2021).
- Government of Pakistan, 2021. COVID-19 Dashboard. Government of Pakistan. <https://covid.gov.pk/stats/pakistan>. (Accessed 17 April 2021).
- Gribble, J., Stevens, L.J., Agostini, M.L., Anderson-Daniels, J., Chappell, J.D., Lu, X., Puijssers, A.J., Routh, A.L., Denison, M.R., 2021. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog.* 17, E1009226.
- Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34, 4121–4123.
- Rambaut, A., Holmes, E.C., O'toole, A., Hill, V., Mccrone, J.T., Ruis, C., Du Plessis, L., Pybus, O.G., 2020. A dynamic nomenclature proposal for sars-cov-2 lineages to assist genomic epidemiology. *Nat Microbiol* 5, 1403–1407.
- Rausch, J.W., Capoferri, A.A., Katusiime, M.G., Patro, S.C., Kearney, M.F., 2020. Low genetic diversity May Be an achilles heel of Sars-Cov-2. *Proc. Natl. Acad. Sci. U. S. A.* 117, 24614–24616.
- Rouchka, E.C., Chariker, J.H., Chung, D., 2020. Variant analysis of 1,040 SARS-CoV-2 genomes. *PLoS One* 15, E0241535.
- Science Media Center, 2021. Expert reaction to cases of variant B.1.617 (The 'Indian Variant') being investigated in the UK, 21st April 2021. https://www.sciencemediacentre.org/expert-reaction-to-cases-of-variant-b-1-617-the-indian-variant-being-investigated-in-the-uk/?Cli_Action=1618957124.823. (Accessed 21 April 2021).