JIDC | THE JOURNAL OF INFECTION IN DEVELOPING COUNTRIES

Coronavirus Pandemic

Sequencing of SARS-CoV-2 in local transmission cases through Oxford Nanopore MinION platform from Karachi Pakistan

Samina Naz Mukry[1,2,3], Shariq Ahmed[1,4], Ali Raza Bukhari[1,4], Aneeta Shahni[2,3], Gul Sufaida[1], Arshi Naz[1,3,4], Tahir Sultan Shamsi[3,5]

[1] Department of Molecular Biology, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan
[2] Department of Transplant Immunology and Applied Microbiology, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan
[3] Department of Post Graduate Studies and Research, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan
[4] Department of Genomics, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan
[5] Department of Clinical Hematology, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan

Abstract

Introduction: The first case of severe acute respiratory syndrome 2 (SARS-CoV-2) was imported to Pakistan in February 2020, since then 8,260 deaths have been witnessed. The virus has been constantly mutating and local transmission cases from different countries vary due to host dependent viral adaptation. Many distinct clusters of variant SARS-CoV-2 have been defined globally. In this study, the epidemiology of SARS-CoV-2 was studied and locally transmitted SARS-CoV-2 isolates from Karachi were sequenced to compared and identify any possible variants.

Methodology: The real time PCR was performed on nasopharyngeal specimen to confirm SARS-CoV-2 with Orf 1ab and E gene as targets. The virus isolates were sequenced through oxford nanopore technology MinION platform. Isolates from the first and second wave of COVID-19 outbreak in Karachi were compared.

Results: The overall positivity rate for PCR was 26.24% with the highest number of positive cases in June. Approximately, 37.45% PCR positive subjects aged between 19-40 years. All the isolates belonged to GH clade and shared missense mutation D614G in spike protein linked to increased transmission rate worldwide. Another spike protein mutation A222V coexisted with D614G in the virus from the second wave of COVID-19.

Conclusions: Based on the present findings it is suggested that the locally transmitted virus from Karachi varies from those reported from other parts of Pakistan. Slight variability was also observed between viruses from the first and second wave. Variability in any potential vaccine target may result in failed trials, therefore information on any local viral variants is always useful for effective vaccine design and/or selection.

Key words: SARS-CoV-2; variants, COVID-19; locally transmitted virus.

J Infect Dev Ctries 2021; 15(10):1376-1383. doi:10.3855/jidc.14900

(Received 12 February 2021 – Accepted 22 April 2021)

Introduction

Coronavirus disease (COVID-19) is a transmissible infectious disease caused by a newly emerged beta coronavirus - SARS-CoV-2 originated as a result of viral spillover from animals [1]. It is a positive sense, enveloped, single stranded RNA virus of genus *Betacoronavirus* [2]. The exact origin of this virus either from the bat, pangolin, or any other mammal is still under debate [1,3]. Metagenomic analysis of SARS-CoV-2 has revealed that it is a distinct virus closely related to SARS CoV. The COVID-19 pandemic started in December 2019 after first reports from Wuhan, China [3,4]. To date, it has affected 188 countries with a death rate of 2.31%. The death rate for SARS-CoV-2 infection is lower than SARS CoV with a higher transmission capability [5]. The global elderly population was most affected with higher mortalities due to acute respiratory distress syndrome (ARDS). The first SARS-CoV-2 genome was sequenced and published in December 2019 by Wang *et al* [6]. The genome of SARS-CoV-2 is ~29.9 kb long with orf1ab at the 5'end and spike protein (S), envelope protein (E) and matrix protein (M) coded at the 3' end. Seven viral accessory proteins are also coded by *ORF3a*, *ORF6*,

ORF7a, *ORF7b*, *ORF8a*, *ORF8b* and *ORF10* genes [7]. The virus has sixteen non-structural proteins (NS1-NS16). The infection initiates with lower respiratory discomfort which progress to pneumonia often causing sudden deaths [4,8]. The virus establishes itself by binding through receptor binding spike protein to angiotensin-converting enzyme 2 (ACE2) receptors in lungs [4]. Data from various studies support virus induced exaggerated immune reaction or cytokine storming leading to extensive damage to ACE2 receptors expressing organs in the host [2].

The rapid spread of the recent outbreak of COVID-19 was due to the dilemma that asymptomatic carriers were capable of transmitting the virus to a healthy human. During SARS CoV outbreak earlier in 2002/2003 variant viruses evolved due to possible transformation events within host [9]. Likewise, SARS-CoV-2 has acquired many mutations with time. A large number of SARS-CoV-2 sequences have been deposited in respective repositories since the beginning of the pandemic [10,11]. Due to relatively lower mutation rates different branches or clades have been defined [12]. The clinical significance of all these clades is yet to be defined. As per Nextstrain an open source project so far five large clades of SARS-CoV-2 (19A, 19B, 20A, 20B and 20C) have been identified all over the world. The earlier strains identified in 2019 belonged to clade 19A and 19B. The clade 19B (GISAID S) differed from the root clade 19A by substitutions of C > T and T > C at positions 8782 and 28144 respectively. The clade 20 A had unique substitutions of C > T at positions 3037 and 14408 along with substitution G > C 28883. The clade 20B originated with consecutive substations 28881 G > A, 28882 G > A, and 28883 G > C whereas 20C has substitutions 1059 C > T and 25563 G > T. The clades 19A and 19B were geographically linked to Asia whereas 20A and 20B were prevalent in Europe suggesting their European origin. Moreover, clade 20C was prevalent in North America in the latter half of 2020. The SARS-CoV-2 sequences submitted from Pakistan corresponds to clades 19A, 19B, 20A and 20B. Rapid whole genome sequencing coupled with prompt data sharing is the key to understand the emergence of variants and geographic epidemiology of SARS-CoV-2 variants during the current pandemic [13]. Clinical correlation of these variants with disease transmission dynamic, treatment responsiveness and fatality rates have also been extensively studied [14]. Among the various sequencing platforms, third generation sequencing of SARS CoV whole genome by Oxford Nanopore MinION technology based sequencing has gained popularity. The advantage of this platform is that long reads of virus genome are obtained and time to data acquisition and analysis is also reduced as compared to other methodologies [15]. The objective of the present study was to evaluate the epidemiology of SARS-CoV-2 in symptomatic and asymptomatic COVID –19 suspected cases with the impact of age and gender. Furthermore, locally transmitted SARS-CoV-2 isolates from Karachi were sequenced to compared and identify any possible variants.

Methodology

This study was conducted at National Institute of Blood Diseases and Bone Marrow Transplantation after approval from NIBD Bioethics committee. A total of 2,065 patients of either gender were enrolled. A detailed proforma was filled in order to obtain the background exposure and brief clinical history of all patients. Nasopharyngeal specimens were collected from suspected COVID-19 patients and transported in VTM to the laboratory for viral RNA extraction using Favorgene viral RNA extraction kit (Cat#: FAVNK 001-2, Fav0rgene Bio Corp, Taiwan). The real time PCR was performed using CE IVD marked Bosphore Anatolia Geneworks kit with orf1ab and E genes as viral targets.

Viral whole genome sequencing

RNA was extracted from six nasopharyngeal samples positive for SARS-CoV-2. All the patients were confirmed local transmission cases from Karachi. Viral RNA extracted from both asymptomatic and symptomatic COVID-19 patients were included. Whole genome sequencing was performed as previously described through ONT MinION platform. Briefly, cDNA was prepared from viral RNA using random primer mix (NEB, Massachusetts, United States), and LunaScript® RT SuperMix Kit (Cat #: E3010G**)**. The cDNA was enriched using Q5 high-fidelity DNA polymerase (Cat #: M0492; NEB, USA) and ARTIC v3 primers (Integrated DNA Technologies (ITD), Belgium). The 400bp amplicons generated were used for library preparation after purification by AMPure XP beads (Agencourt Beckman Coulter™ A63881) and quantification with a Qubit fluorometer using Qubit™ RNA BR Assay kit (Cat #: Q10210, ThermoFisher Scientific, Massachusetts, United States). For sequencing, SQK-LSK104 ligation kit was used and samples were loaded as per standard protocol on a MinION MK1B instrument. All the samples were multiplexed in one run after barcoding. The samples were run for 72 hours on the R9 flow cell. The acquired

data were aligned in real time with SARS-CoV-2 reference sequence (NC_045512.2) using RAMPART (MiniKnow). Files in SAM/BAM format were accessed by SAM tools v. 1.9-11.

Variant calling and phylogenetic profiling

Variant calling was performed using BCF tools v. 1.9. Further variant annotation was done by ANNOVAR. The consensus sequences were generated by mapping the variants to the reference genomes using BCF tools followed by submission to the GISAID database and NCBI. The initial phylogenetic analysis was performed using 570 genome sequences; including all available Pakistani SARS-CoV-2 genomes; retrieved from the NCBI SARS-CoV-2 resources (GISAID, epicov^{TM}). The fast alignment was performed using MAFFT. Moreover, IQ-TREE v. 2.1.2 through pipeline adopted by Augur [16] was used to infer maximum-likelihood trees. The generated tree was visualized using FigTree 1.4.3 [17]. Viral clades were identified by nucleotide or amino acid substitutions and Nextstrain nomenclature [18] was followed for matching. The robustness of individual nodes was statistically determined by comparing with 1,000 bootstraps replicate. The viral lineage was determined by using Phylogenetic Assignment of Named Global Outbreak LINeages tool [19].

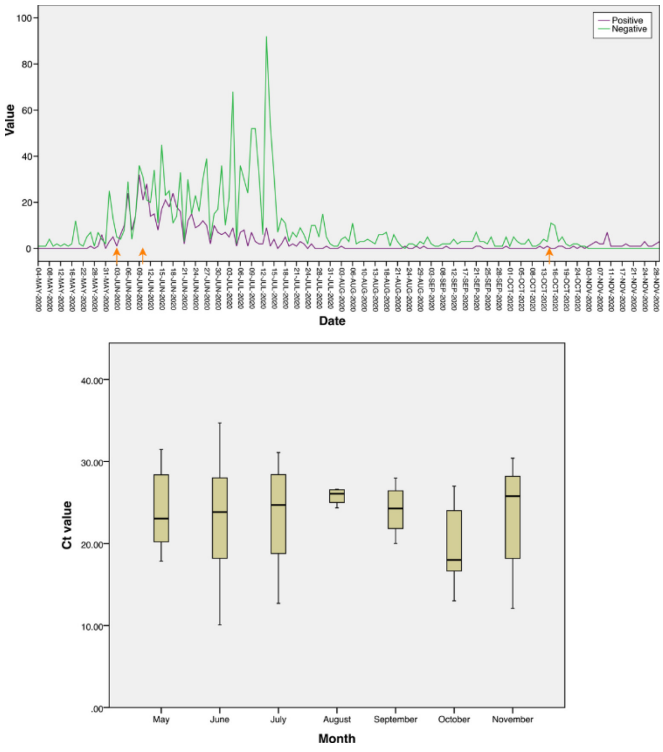
Statistical analysis

Median Ct values per month for Orf1ab gene were recorded from May to November 2020. For statistical analyses and graphs SPSS version 22 was used.

Results

A total of 2065 SARS-CoV-2 suspected nasopharyngeal samples were tested by PCR at COVID-19 Lab. of NIBD during May to November 2020. The overall positivity rate for PCR was 26.24%. The highest number of positive cases with increased viral load (lower Ct values) was observed during the month of June (Figure 1a). Moreover, the Ct value for SARS-CoV-2 ranged between 10.8 to 34.32 in June with a median Ct value of 24.2 (Figure 1b). A large number of patients were negative for SARS-CoV-2 with COVID-19 like symptoms caused by probably some other viral or bacterial infection. The common

Figure 1. Detailed overview of SARS-CoV-2 suspected samples analyzed at National Institute of Blood Diseases and Bone Marrow Transplantation.



Nasopharyngeal swabs were tested for the presence of SARS-CoV-2 RNA using commercial kit. X axis (horizontal) shows dates from 4th May to 30th November 2020 when samples were taken. A) Graphical summary of tested specimens per day. The green line shows samples negative SARS-CoV-2 per day and the blue line shows positive samples. The dots indicate total samples for each day. The red arrow indicates the date of collection for samples 1= NIBD01-PAK-KHI, 2=NIBD02-PAK-KHI, 3=NIBD03-PAK-KHI, NIBD04 selected for SARS-CoV-2 whole genome sequencing. B) The box and whiskers graph where Ct median value with SARS-CoV-2 specific Orf1ab as target is indicated by line and the box is extended from 25th to the 75th percentiles. Lowest median Ct value 20.21 was observed in October.

Table 1. Detailed genomic features of four SARS CoV2 viruses.

Sample	Genbank Accession Number	GISAID Accession No.	Lineage	No of reads	genome Size	GC content	coverage X	Depth of Coverage	Mapped reads
NIBD01-PAK-KHI	MW403500	EPI_ISL_708839	B.1.36.6	193131	29603	39%	100	2976	98.57%
NIBD02-PAK-KHI	MW400961	EPI_ISL_709540	B.1.36.6	240238	29836	39%	100	2976	98.58%
NIBD03-Pak-Khi	MW411960	EPI_ISL_709544	B.1.160.	232888	29594	40%	100	3653	98.57%
NIBD04-Pak-Khi	MW411961	EPI_ISL_709542	B.1.36	197900	29519	40%	100	3161	86.77%

indication was weakness for PCR positive and PCR negative groups. The frequency of SARS-CoV-2 positive males (27.5%) was slightly higher than females (26.28%). Approximately 37.45% PCR positive subjects aged between 19-40 years.

Viral whole genome sequencing

Six samples with low Ct values < 20 (range: 10.08-19.69) were selected. Two samples were not processed further due to failed QC after cDNA enrichment step. Of the 4 remaining samples, two were isolates from asymptomatic patients and one from mild and moderate COVID-19 –patients each. The age range of selected subjects was 22-58 years. Viral genome sequences from these confirmed cases of local transmission of SARS-CoV-2 from Karachi during June (peak of first wave) and October (initial phase of second wave) were obtained through MinIon ONT platform. The sequencing details with GISAID accession numbers are listed in Table 1. The genome size obtained was 29903 with depth of coverage between 2976-3653 (Table 1).

Phylogenetic profiling

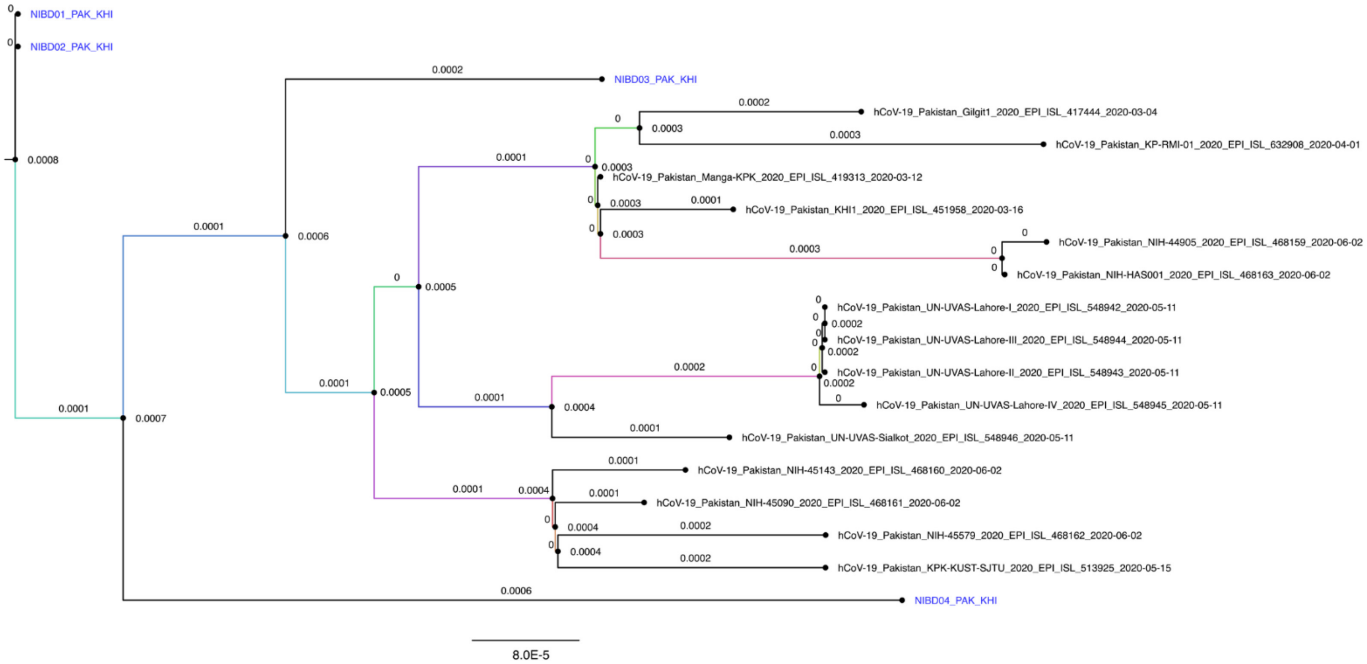
The whole genome sequences obtained were aligned with the reference genome and 570 global sequences. The strains NIBD 01-PAK-KHI and NIBD 02-PAK-KHI clustered with variants from Bangladesh and India with descendent predominantly from Saudi Arabia, India and England. New nodes were defined for NIBD 03-PAK-KHI and NIBD 04-PAK-KHI with the divergence of 14 and 27 respectively. Both these were clustered with separate sequences from New Zealand having presumed ancestral connection with isolates from India, Saudi Arabia and England.

The phylogenetic relatedness with previously submitted sequences from Pakistan, India, Saudi Arabia, Netherland, England, New Zealand was observe with a pair-wise mean genetic distance of 0.011.

Mutational analysis

The mutational analysis revealed the presence of 21 synonymous mutations, 15 non-synonymous mutations and 2 non-frame shift substitutions altogether, spanning in 5'UTR, spike, orf1ab, orf1a, orf3a, orf7a, orf 8, orf10, N and M protein genes (Table 2). All the four sequences were grouped in clade 20A (GISAID: GH) since characteristic mutation in 5'UTR 241 C>T and nonsynonymous SNV i.e. 3037 C > T, 23403 A > G (S-D614G), 25563 G > T (Q57H) were observed. As per GISAID database, four previously submitted sequences from Pakistan also belonged to GH clade but sub-lineage signatures vary with 32 additional mutations observed in the sequences obtained during the present study from local transmission cases of Karachi (Table 2).

Figure 2. Phylogenetic tree for locally isolated SARS CoV 2 viruses from Pakistan.



The locally transmitted SARS-CoV-2 from Karachi clustered distinctly with viruses from Kohat and Islamabad. The root for the tree is reference sequence from China NC_045512.2.

Table 2. List of single nucleotide variants in genome of locally transmitted SARS CoV2 from Karachi.

Sample Ids	Mutation	gene	gene location	amino acid change	Type of Single nucleotide variant
1,2,3,4	*241C>T	5′UTR	untranslated region	-	NA
3	458G>A	orf1ab	nsp1	E65K	*nonsynonymous_SNV
3	1351A>G,	orf1ab	nsp2	p.Q362Q	synonymous_SNV
3	1451A>G	orf1ab	nsp2	I396V/I216V	nonsynonymous_SNV
4	1722C>T	orf1a	nsp2	A486V	nonsynonymous_SNV
1,2,3,4	*3037C>T	Orf1ab	nsp3	F924F	*synonymous_SNV
1	4126G>A	orf1ab	nsp3	V1287V	synonymous_SNV
1	5482C>T	orf1ab	nsp3	A1739A	synonymous_SNV
4	8512A>G,	orf1ab	nsp3	Q2749Q	synonymous_SNV
4	9877T>C	orf1ab	nsp4	Y3204Y	synonymous_SNV
4	10985A>G	orf1ab		R3574G	nonsynonymous_SNV
4	11222G>T	orf1ab	nsp6	V3653F	nonsynonymous_SNV
4	12616C>T,	orf1ab	nsp8	p.D4117D	synonymous_SNV
1,2,3,4	*14408C>T	ORF1ab	NSP12	P323L	*nonframeshift_substitution
4	15738C>T	orf1ab	NSP12	F5158F	synonymous_SNV
4	16716C>T	orf1ab	helicase	D5484D	synonymous_SNV
3	16915C>T	orf1ab	Helicase	L5551L	synonymous_SNV
2	17502C>T	orf1ab		F5746F	synonymous_SNV
1,2,3,4	18877C>T	ORF1ab	3'-to-5' exonuclease	L6205L	synonymous_SNV
4	21868A>T	Spike	nsp11	R102S	nonsynonymous_SNV
4	21868A>T	Spike	surface glycoprotein	R102S	nonsynonymous_SNV
4	**22227C>T	Spike	surface glycoprotein	A222V	**nonsynonymous_SNV
1	22444C>T	spike	surface glycoprotein	D294D	synonymous_SNV
1,2,3,4	*23403A>G	spike	surface glycoprotein	D614G	*nonsynonmous
3	24202T>C	S	surface glycoprotein	G880G	synonymous_SNV
4	25463C>T	ORF3a	NS3	T24I	nonsynonymous_SNV
1,2,3,4	*25563G>T	ORF3a	NS3	Q57H	*nonsynonymous_SNV
4	25916C>T	ORF3a	NS3	T175I	nonsynonymous_SNV
3	26534C>T	M	membrane glycoprotein	S4S	synonymous_SNV
1,2,3	26735C>T	M	membrane glycoprotein	Y71Y	synonymous_SNV
4	26735C>T	M	membrane glycoprotein	T77T	synonymous_SNV
4	27561G>T	ORF7a	_	L56L	synonymous_SNV/Nonframeshift substitution
4	*28253C>T	ORF8	_	F120F/I120S	synonymous_SNV
1	28854C>T	N	nucleocapsid phosphoprotein	S194L	nonsynonymous_SNV
4	28854C>T	N	nucleocapsid phosphoprotein	S194L	nonsynonymous_SNV
3	28899G>T	N	nucleocapsid phosphoprotein	R209I	nonsynonymous_SNV
4	28940C>T	N	nucleocapsid phosphoprotein	L223L	synonymous_SNV
4	29095C>T	N	nucleocapsid phosphoprotein	F274F	synonymous_SNV
4	29628G>A	ORF10	NS3	R24H	nonsynonymous_SNV

1: NIBD01-PAK-KHI; 2: NIBD02-PAK-KHI; 3: NIBD03-PAK-KHI; 4: NIBD04; * Clade defining variants observed in other GH clade variants from Pakistan; **Characteristic mutation of GV clade.

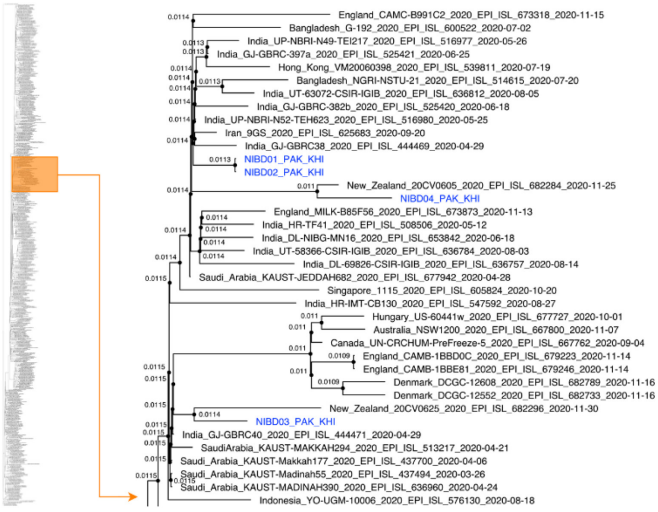
Hence, all the genome sequences were unrelated to the previously reported cluster of SARS-CoV-2 from Pakistan except for four isolates; 3 from Islamabad and 1 from Kohat belonging to GH clade lineage B (Table 1; Figure 3). The NIBD4-PAK-KHI obtained from a health care worker varied from the other isolates with the highest number of mutations (Table 2). An additional clade GV specific nonsynonymous variant 22227 C > T (A222V) in spike protein was also present in NIBD4-PAK-KHI along with D614G mutation. About 7% of all GISAID sequences belonged to GV clade which is characterized by the presence of A222V SNV. The virus NIBD4-PAK-KHI is the first variant of GH clade harboring 22227 C > T isolated in October from Asia. Close but divergent sequence homology was detected with a variant virus from New Zealand collected in November 2020 (GISAID accession no: EPI_ISL_682284; Figure 3).

Discussion

Karachi is the most populous major metropolitan city of Pakistan. The first imported case of COVID-19, was detected in Karachi in February 2020, in a returning traveler from Iran. After that the city became the worst-hit area of Pakistan during the first wave of COVID-19. This study is one of the initial reports on the epidemiological and molecular landscape of SARS-CoV-2 from Karachi. Based on the retrospective data from NIBD the PCR positivity rate for SARS-CoV-2 was 24.2% in June (the peak of first wave) which dropped to 5% during September. The positive cases per day increased in October (Figure 1) with a lowest median Ct value observed in October as compared to other months. The increasing trend in PCR positivity during this period correlated well with the regional increase in COVID-19 cases in Sindh. In order to contain SARS-CoV-2 transmission and also to reduce the possible burden on the striving national healthcare system, the government of Pakistan officially imposed a smart lockdown declaring a second wave of COVID-19 with increased cases on 28th October 2020. In order to evaluate any possible genetic variability in community transmitted SARS-CoV-2 in Karachi, four confirmed indigenous SARS-CoV-2 viruses were randomly selected from the peak of first wave and initial phase of second wave of COVID-19. The genome sequencing was performed through ONT MinION platform. The ONT is claimed to be the most accurate, sensitive and quick method with reduced PCR based bias for detecting minute changes in viral genome. Samples are barcoded to facilitate multiplexing and the tiling generates precise long stretches of genome which could be read easily through defined pipelines thereby decreasing the time for post-acquisition analysis of SARS-CoV-2 genome [15]. To the best of our knowledge, this is the first study from Pakistan on SARS-CoV-2 genome exploiting the potential of this third generation sequencing method.

The evolution of SARS-CoV-2 is nonrandom and human host dependent variants are evolving. With over 272,829 virus sequence depositions in GISAID Initiative (epicov.org) alone [20]. SARS-CoV-2 proteins are heterogeneous with a large number of variable amino acid substitutions with either no or significant impact on viral transmission and transcription within human host [21]. Many prevalent mutations have been defined for SARS-CoV-2 with signature hotspot mutation for each distinct clade [22]. The lineages for studied viruses were defined through Pangoline pipeline and all were placed in lineage B1 with sub-lineage differences listed in Table 1. The 23403 A > G is a widely documented hotspot mutation in spike protein that replaces aspartic acid with glycine at position 614 altering the viral antigenic properties [23]. Variants with D614G evade initial immune recognition by the host resulting in production of autoantibodies and facilitate higher- transmission, infectivity and case fatality rate (CFR). All the sequenced viruses of the current study had this mutation along with Q57H in orf3a gene grouping them in Next strain clade 20A (GISAID: GH; Table 2). The coexisting Q57H has previously been reported to reduce the virulence of D614G therefore the prevalence of the GH clade in Pakistan may be a probable reason for low mortality rate (2.1%) for COVID-19 cases

Figure 3. Geo-epidemiological profile of SARS-CoV-2 viruses from Karachi. The out group for the tree is Wuhan reference sequence NC_045512.2.



during July to September. Furthermore, the Q57H amino acid substitution causes truncation of orf3b gene *via* introduction of a stop codon at amino acid 13 giving rise to full length orf3b deficient variants [24,25]. These variants were prevalent in Asian and North American countries, including Saudi Arabia, Indonesia, South Korea, Israel, Egypt, USA and Colombia [5,26,27]. All Pakistani isolates sequenced between June to November 2020 belonged to GH clade with Q57H substitution (Figure 3). The orf3b is a potential serological target for most vaccines; hence, this serological difference should be taken into account while selecting a vaccine for SARS-CoV-2 in Pakistan. Another prevalent mutation in Pakistani GH isolates is non-frameshift substitution P323L in NSP12 which was linked to higher severity [24]. Almost a 100% coexistence of D614G, P323L and C241T has been reported which is in line with the present observations. This coexistence positively favors viral replication, infectivity, transmission and manipulation of host machinery [28]. Of the 39 mutations observed in the SARS-CoV-2 genomes sequenced during the present study; 33 SNVs were not observed in previously submitted sequences from Pakistan defining a separate local cluster for SARS-CoV-2 virus accumulating within Karachi (Figure 2). It may be because of the host driven genetic drift within the viral genome of locally transmitted SARS-CoV-2 virus [29].

Spike A222V mutation was reported from Spain during spring 2020 and clade GV (20A EU1) of SARS-CoV-2 was defined with increased infectivity and fatality potential [30]. Many mutations overlap between G clade. The global coexistence of A222V mutation within GH clade viruses is rare (0.0015%; GISAID Initiative (epicov.org) with total of 84 viruses. The NIBD04 PAK-KHI carried this mutation and is the first sequence from Asia with this unique coexistence sharing ancestral origin from Europe. Within Europe and USA; the A222V is associated with newer waves of COVID-19 with increase viral loads, similarly the virus from the initial phase of second wave of COVID-19 in Karachi was presented with it. It can be assumed that the median viral load corresponding with a lower median Ct value may be an indirect indication of accumulation of this virus in Karachi which needs further confirmation *via* sequencing of virus from the second wave. Another possible reason for this variance in NIBD04 PAK-KHI could be co-infection with two variants of SARS-CoV-2. The co-infection has been reported with a low prevalence rate of 1.5% in healthcare workers globally. In either case, caution is warranted since infectivity potential for this variant is higher throughout Europe.

Conclusions

Hence it can be concluded that the second wave of COVID-19 may not be clinically distinct but host driven genetic evolution of virus may impact its infectivity and CFR. Future comparative genomic studies with a substantial number of SARS-CoV-2 isolates from first and second waves are suggested in order to understand the possible evolutionary origin, genomic variability and trajectory of anticipated futures waves.

Acknowledgements

The authors are thankful to Scientechnic for providing the MinION MK1B instrument and timely supply of all ordered consumable for this project during the pandemic.

References

1. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579: 270–273.
2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395: 565–574.
3. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou J-J, Li N, Guo Y, Li X, Shen X (2020) Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature 583: 286–289.
4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395: 497–506.
5. Mercatelli D, Giorgi FM (2020) Geographic and genomic distribution of SARS-CoV-2 mutations. Front Microbiol 11: 1800.
6. Wang C, Horby PW, Hayden FG, Gao GF (2020) A novel coronavirus outbreak of global health concern. Lancet 395: 470–473.
7. Yoshimoto FK (2020) The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. Protein J 39: 198–216.
8. Fung S-Y, Yuen K-S, Ye Z-W, Chan C-P, Jin D-Y (2020) A tug-of-war between severe acute respiratory syndrome coronavirus 2 and host antiviral defence: lessons from other pathogenic viruses. Emerg Microbes Infect 9: 558–570.
9. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y (2020) A new coronavirus associated with human respiratory disease in China. Nature 579: 265–269.
10. Rahimi A, Mirzazadeh A, Tavakolpour S (2020) Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. Genomics 113: 1221-1232.

11. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CCS, Boshier FAT (2020) Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol 83: 104351.
12. Hodcroft EB, Hadfield J, Neher RA Bedford T (2020) Year-letter genetic clade naming for SARS-CoV-2 on nextstrain.org. Available: https://nextstrain.org/blog/2020-06-02-SARSCoV2-clade-naming. Accessed: 20 Decemebr 2020.
13. Hartley PD, Tillett RL, AuCoin DP, Sevinsky JR, Xu Y, Gorzalski A, Pandori M, Buttery E, Hansen H, Picker MA, Rossetto CC, Verma SC (2020) Genomic surveillance of Nevada patients revealed prevalence of unique SARS-CoV-2 variants bearing mutations in the RdRp gene. MedRxiv Preprint. 2020.08.21.20178863.
14. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B (2020) Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182: 812–827.
15. Moore SC, Penrice-Randal R, Alruwaili M, Dong X, Pullan ST, Carter D, Bewley K, Zhao Q, Sun Y, Hartley C (2020) Amplicon based MinION sequencing of SARS-CoV-2 and metagenomic characterisation of nasopharyngeal swabs from patients with COVID-19. MedRxiv Preprint. 2020.03.05.20032011.
16. Nextstrain (2020) Nextstrain/Augur: pipeline components for real-time phylodynamic analysis. Available: https://github.com/nextstrain/augur. Accessed: 26 December 2020.
17. Rambaut A (2009) FigTree, version 1.4.3. Available: http://tree.bio.ed.ac.uk/software/figtree. Accessed: 26 December 2020.
18. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA (2018) Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34: 4121–4123.
19. GitHub (2020) cov-lineages/pangolin: software package for assigning SARS-CoV-2 genome sequences to global lineages. Available: https://github.com/cov-lineages/pangolin. Accessed: 26 December 2020.
20. GISAID (2020) Clade and lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses Available: https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/. Accessed: 26 December 2020.
21. Yao H, Lu X, Chen Q, Xu K, Chen Y, Cheng L, Liu F, Wu Z, Wu H, Jin C, Zheng M, Wu N, Jiang C, Li L (2020) Patient-derived mutations impact pathogenicity of SARS-CoV-2. MedRxiv Preorint 2020.04.14.20060160.
22. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC (2020) Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med 18: 179.
23. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L (2020) The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell 182: 1284–1294.
24. Bianchi M, Borsetti A, Ciccozzi M, Pascarella S (2021) SARS-Cov-2 ORF3a: mutability and function. Int J Biol Macromol 170: 820–826.
25. Lam JY, Yuen CK, Ip JD, Wong WM, To KKW, Yuen KY, Kok KH (2020) Loss of orf3b in the circulating SARS-CoV-2 strains. Emerg Microbes Infect 9: 2685–2696.
26. Laamarti M, Alouane T, Kartti S, Chemao-Elfihri MW, Hakmi M, Essabbar A, Laamarti M, Hlali H, Bendani H, Boumajdi N (2020) Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations. PLoS One 15: e0240345.
27. Hodcroft EB, Zuber M, Nadeau S, Vaughan TG, Crawford KHD, Althaus CL, Reichmuth ML, Bowen JE, Walls AC, Corti D, Bloom JD, Veesler D, Mateo D, Hernando A, Comas I, González Candelas F; SeqCOVID-SPAIN consortium, Stadler T, Neher RA (2020) Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. MedRxiv Preprint. 2020.10.25.20219063.
28. Kamikubo Y, Takahashi A (2020) Epidemic trends of SARS-CoV-2 modulated by economic activity, ethnicity, and vaccination. Cambridge Open Engag Preprint coe-2021-xfgln.
29. Liu S, Shen J, Fang S, Li K, Liu J, Yang L, Hu C-D, Wan J (2020) Genetic spectrum and distinct evolution patterns of SARS-CoV-2. Front Microbiol 11: 2390.
30. Bartolini B, Rueca M, Gruber CEM, Messina F, Giombini E, Ippolito G, Capobianchi MR, Di Caro A (2020) The newly introduced SARS-CoV-2 variant A222V is rapidly spreading in Lazio region, Italy. MedRxiv Preprint. 2020.11.28.20237016.

Corresponding author

Dr. Samina Naz Mukry, PhD, Associate Professor,
Department of Post Graduate Studies and Research, Department of Transplant Immunology and Applied Microbiology, Department of Molecular Biology, National Institute of Blood Diseases Bone Marrow Transplantation, ST 2/A Block 17 Gulshan-e-Iqbal KDA Scheme 24, Karachi, Pakistan. UAN:111-120-140
Phone: 92-21-34824250-3 Ext:405,
Fax: 92-21-34821504
Email: smukry.nibd@gmail.com; smukry@gmail.com; saminanaz.mukry@nibd.edu.pk

Conflict of interests: No conflict of interests is declared.